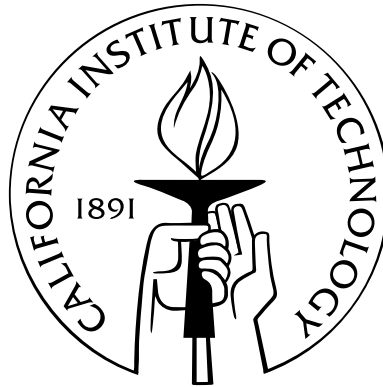


# Demographic Studies of Extrasolar Planets

Thesis by  
Timothy Morton

In Partial Fulfillment of the Requirements  
for the Degree of  
Doctor of Philosophy



California Institute of Technology  
Pasadena, California

2014  
(Submitted October 28, 2013)



*To V&A, my beloved family.*

# Acknowledgements

I am very grateful to all those who have enabled me to have the extraordinary privilege to pursue this research, which skirts the boundaries of human knowledge of our place in the cosmos. This long list of people begins, of course, with my parents Dan and Roberta, whose every decision was geared toward giving me and my sisters Deborah and Rebecca (whom I also acknowledge for being generally awesome at life and helping me in countless ways throughout the years) the best educational opportunities we could possibly have. The story continues with all my teachers at Columbus Academy who taught, encouraged, and challenged me over the years, especially Mrs. Corrigan (6th and 7th grade science), Mr. Foltz (8th grade science), Mr. Booth (math), Mrs. Haque (physics), Mrs. Hogan (English/writing), and Mr. Daubenmier (computer science/yearbook/life). From my Harvard years, I thank Howard Georgi for a phenomenal Physics 16 class that directly encouraged me to continue to study physics, and all my physics/astronomy friends (especially Jon, Erika, and Ben) without whom I never would have been able to handle all the problem sets, let alone the trip to the Quad every Friday for tutorial. Thanks to the OSU REU program that gave me my first taste of “research” (colored pencils and all), and to Kris Stanek, Jenny Greene, and Patrick Slane, who mentored me through my first real projects, and apparently thought highly enough of me to recommend me well to the next stage. Also (a bit out of order), special thanks to Ramesh for a fantastically taught Astro 145 course. In the Caltech astronomy department, I gratefully acknowledge everyone who enables it to be the great hub of leading-edge research that it is, as well as all those who had a hand in my statistically improbable admission to the graduate program. I’m thankful also to my classmates (Drew, Elisabeth, Gwen, Krzysztof, Mike, Sam, Tucker, and Vera) and officemates through the years (Vera, Mike, Adam, Matthew, Krzysztof) who have been fellow travelers, entertainers, co-endurers, and resources during this six-year journey. And most especially I would like to extend deep thanks to my adviser John Johnson for his willingness to take me on as a third-year student his first year at Caltech. Without his unrelenting positivity, his continual encouragement, his boundless energy and enthusiasm, his selfless honesty, his foresighted guidance, his birthing of the fantastic ExoLab (thanks also to the postdocs therein for their friendship/encouragement/guidance—Justin, Sasha, Phil, Jon, Leslie, Avi), and even his measured criticisms (not to mention his ball-handling and dime-dropping skills), I would never have been in the position to even begin this quest, let alone finish it. It’s truly an honor to be his first Ph.D. student.

I also acknowledge the many useful conversations and interactions with astronomers outside of Caltech that helped inspire and improve my research and understanding of statistics, including (but not limited to) Josh Winn, Dan Fabrycky, Ed Turner, Eric Agol, Geoff Marcy, Andrew Howard, Willie Torres, Scott Gaudi, Steve Bryson, Jason Eastman, Erik Petigura, David Kipping, and those involved with the Penn State Astrostatistics Summer School. Special thanks also to Josh Winn for being the PI on the grant that supported my research the last three years.

In addition, I would like to extend thanks to those outside Caltech who made my time living in the LA area so vibrant and memorable, especially all the choral musicians (especially Jay, Sarah, Christian, and Jim) and friends and mentors from St. James' (nota bene Trent, Todd, Ben, Susanne, the Woods, Jon, and Anne). Without communities like this to share our lives with, what good is all the exoplanet knowledge in the Universe?

And finally, I thank Vera—my classmate, officemate, friend, roommate, wife, Doctor, mother-of-my-son, and teammate-in-everything-forever—specifically for her constant support during “thesis tiiiime,” and in general for helping me always to remember what is most important in life, and for making that life so wonderful. Without a family like this, what good is all the exoplanet *and* cosmology knowledge in the Universe?

# Abstract

Uncovering the demographics of extrasolar planets is crucial to understanding the processes of their formation and evolution. In this thesis, we present four studies that contribute to this end, three of which relate to NASA’s *Kepler* mission, which has revolutionized the field of exoplanets in the last few years.

In the pre-*Kepler* study, we investigate a sample of exoplanet spin-orbit measurements—measurements of the inclination of a planet’s orbit relative to the spin axis of its host star—to determine whether a dominant planet migration channel can be identified, and at what confidence. Applying methods of Bayesian model comparison to distinguish between the predictions of several different migration models, we find that the data strongly favor a two-mode migration scenario combining planet-planet scattering and disk migration over a single-mode Kozai migration scenario. While we test only the predictions of particular Kozai and scattering migration models in this work, these methods may be used to test the predictions of any other spin-orbit misaligning mechanism.

We then present two studies addressing astrophysical false positives in *Kepler* data. The *Kepler* mission has identified thousands of transiting planet candidates, and only relatively few have yet been dynamically confirmed as *bona fide* planets, with only a handful more even conceivably amenable to future dynamical confirmation. As a result, the ability to draw detailed conclusions about the diversity of exoplanet systems from *Kepler* detections relies critically on understanding the probability that any individual candidate might be a false positive. We show that a typical *a priori* false positive probability for a well-vetted *Kepler* candidate is only about 5-10%, enabling confidence in demographic studies that treat candidates as true planets. We also present a detailed procedure that can be used to securely and efficiently validate any individual transit candidate using detailed information of the signal’s shape as well as follow-up observations, if available.

Finally, we calculate an empirical, non-parametric estimate of the shape of the radius distribution of small planets with periods less than 90 days orbiting cool ( $T_{\text{eff}} < 4000$  K) dwarf stars in the *Kepler* catalog. This effort reveals several notable features of the distribution, in particular a maximum in the radius function around 1-1.25 Earth radii and a steep drop-off in the distribution larger than 2 Earth radii. Even more importantly, the methods presented in this work can be applied to a broader subsample of *Kepler* targets to understand how the radius function of planets changes across different types of host stars.

# Contents

<b>Acknowledgements</b>	<b>iv</b>
<b>Abstract</b>	<b>vi</b>
<b>List of Figures</b>	<b>x</b>
<b>List of Tables</b>	<b>xxi</b>
<b>I Introduction</b>	<b>1</b>
<b>1 The History and Goals of Exoplanet Demographics</b>	<b>3</b>
<b>2 Transiting Planets and the <i>Kepler</i> Revolution</b>	<b>9</b>
2.1 False Positives in Transit Surveys . . . . .	10
2.2 The <i>Kepler</i> Mission . . . . .	11
<b>3 Overview of Bayesian Inference</b>	<b>16</b>
3.1 Parameter Estimation . . . . .	17
3.2 Model Selection . . . . .	20
3.3 Final Thoughts on Bayesian Inference . . . . .	22
<b>II Original Research</b>	<b>24</b>
<b>4 Discerning Exoplanet Migration Models Using Spin-Orbit Measurements</b>	<b>25</b>
4.1 Introduction . . . . .	25
4.2 The Models . . . . .	26
4.3 Data Analysis . . . . .	29
4.3.1 Which misalignment mechanism is preferred? . . . . .	29
4.3.2 Confidence Assessment . . . . .	30
4.4 Multiple migration channels? . . . . .	31

4.4.1	What fraction of systems are misaligned? . . . . .	32
4.4.2	Two-mode model selection and confidence assessment . . . . .	33
4.4.3	One channel or two channels? . . . . .	36
4.5	Hot Stars . . . . .	36
4.6	How many hot star RM observations are needed? . . . . .	37
4.7	Discussion and Conclusions . . . . .	40
<b>5</b>	<b>On the Low False Positive Probabilities of Kepler Planet Candidates</b>	<b>43</b>
5.1	Introduction . . . . .	43
5.2	Basic Bayesian Framework . . . . .	45
5.2.1	Blended Binaries . . . . .	46
5.2.1.1	Probability of a blend . . . . .	47
5.2.1.2	Probability of an appropriate eclipsing binary . . . . .	49
5.2.2	Hierarchical Triples . . . . .	51
5.2.3	Basic Framework: Summary and Discussion . . . . .	53
5.3	Detailed Framework: Considering Transit Depth . . . . .	55
5.4	Results . . . . .	58
5.4.1	General . . . . .	58
5.4.2	Application to Kepler Candidates . . . . .	60
5.5	Discussion: Relationship to “BLENDER” . . . . .	61
5.6	Caveats and Conclusions . . . . .	62
5.7	Appendix: Blended Planets . . . . .	64
<b>6</b>	<b>An Efficient Automated Validation Procedure for Exoplanet Transit Candidates</b>	<b>66</b>
6.1	Introduction . . . . .	66
6.2	Framework . . . . .	68
6.3	Procedure . . . . .	69
6.3.1	Population Simulations . . . . .	69
6.3.2	Priors . . . . .	71
6.3.3	Likelihoods . . . . .	73
6.3.4	Final Calculation . . . . .	75
6.4	Tests . . . . .	77
6.4.1	Kepler-19b . . . . .	77
6.4.2	Kepler planet test sample . . . . .	80
6.4.3	Known false positives . . . . .	80
6.4.4	Testing assumptions and simulation variance . . . . .	81
6.5	Relation to Morton & Johnson (2011) . . . . .	82



6.6	Conclusions . . . . .	83
<b>7</b>	<b>The Radius Distribution of Planets Around Cool Stars</b>	<b>89</b>
7.1	Introduction . . . . .	89
7.2	Formalism . . . . .	91
7.2.1	Occurrence Calculations . . . . .	91
7.2.2	Estimating the Radius Distribution Function . . . . .	94
7.3	Calculating the Cool KOI Radius Function . . . . .	95
7.3.1	Period distribution . . . . .	96
7.3.2	SNR distribution . . . . .	96
7.3.3	Radius Distribution . . . . .	98
7.4	Results . . . . .	99
7.5	Exploring Assumptions . . . . .	101
7.5.1	Period Distribution Assumption . . . . .	102
7.5.2	SNR Ramp Assumption . . . . .	103
7.6	Conclusions . . . . .	103
<b>8</b>	<b>Conclusion</b>	<b>105</b>

# List of Figures

- 1.1    Unfiltered distributions of measured exoplanet quantities from the Exoplanet Orbit Database at `exoplanets.org`, demonstrating both the power and the danger of inspecting simple scatter plots in the service of exoplanet demographics. The left-hand panel shows very clearly that many exoplanets likely have more violent dynamical histories than the circular planets of our solar system. The right-hand panel shows the minimum masses ( $M \sin(i)$ ) and periods of all known planets; however, proper interpretation of this plot is more difficult because of the various different detection biases at work. Larger and shorter-period planets are easier to detect, and especially transit surveys (which are represented here, in addition to radial velocity surveys) are very heavily biased toward detecting large and close-in planets. However, the increase in the density of detected giant planets beyond orbital periods of  $>200$  days is a true feature of this distribution. . . . . 4
- 1.2    Two different views of the planet-metallicity relation. The left panel (Santos et al. 2004) shows the distribution of the metallicity of stars with known planets c. 2004 (red-hashed histogram) compared to a volume-limited comparison sample of stars (open histogram), indicating that planet-hosting stars are preferentially metal-rich. The right-hand panel (Fischer & Valenti 2005) quantifies this same relationship more precisely, measuring the occurrence rate (number of detections divided by total number of stars surveyed) in different metallicity bins. . . . . 5
- 1.3    Results of the *NASA-UC eta-Earth* survey (Howard et al. 2010), which was a high-precision RV survey of 166 FGK dwarf stars intended to measure the occurrence rate of planets down to much lower masses than had previously been done, showing that “super-Earth” planets are much more common than giant planets. The key to accomplishing this was to precisely quantify the fractional completeness of the survey, enabling correction for “missed planets.” . . . . 6
- 1.4    The history of exoplanet discovery. Black crosses indicate planets detected in RV surveys, blue dots indicate planets detected in transit surveys, and red dots indicate planets detected by the *Kepler* mission. Despite the evident advances in RV instrumentation, the vast majority of planets with measured velocity amplitudes remain giant planets (Jupiter’s perturbation on the Sun has a semiamplitude of about 10 m/s). . . . . 7

2.1	All exoplanets currently known with both masses and radii measured. <i>Kepler</i> discoveries (red points) dominate the low-mass, small-radius end of this distribution. . . . .	13
2.2	The difficulty following up <i>Kepler</i> candidates with the same strategies that have been used to follow up ground-based transit candidates. The vast majority of <i>Kepler</i> candidates are either too small to detect via RV measurements or are around host stars that make precise measurements too time-consuming. . . . .	14
3.1	Demonstration of Bayesian parameter estimation. The left panel shows a linear model being tested against observed data. The total likelihood of the data $\mathcal{L}$ is the product of all the individual likelihoods $\ell_i$ , each of which is determined by evaluating a likelihood function (inset, red-dotted) at the value of the observed data point. In this example each individual likelihood function is a normalized Gaussian centered at the model value predicted for that time, whose width is determined by the observational uncertainty for that point. The right panel shows the posterior probability distribution of the model parameters $a$ and $b$ , along with the marginalized distribution for each. The contours enclose 68%, 95%, and 99% of the total probability. . . . .	19
4.1	Probability distributions for the true ( $\psi$ ; top panel) and projected ( $\lambda$ ; bottom panel) spin-orbit angles that our two misalignment mechanisms produce. The KCTF $\psi$ distribution is taken from the simulations of final inclinations of planets migrating through Kozai cycles + tidal damping from Fabrycky & Tremaine (2007), and the planet-planet scattering (PSTF) prediction is taken from the simulations of Nagasawa et al. (2008). We use Monte Carlo simulations to project each $\psi$ distribution to create the $\lambda$ distributions. Note that non-zero but small $\lambda$ values are preferred by the KCTF model, while large values ( $\lambda > 60^\circ$ ) are preferred by the PSTF model. The bottom panel shows the distribution of the 32 measured $\lambda$ values. . . . .	28
4.2	The relationship between our model selection statistic $\mathcal{R}$ , or the logarithm of the likelihood ratio, and a model selection confidence. We randomly draw 5000 sets of 32 $\lambda$ values from each model and measure $\mathcal{R}$ for each of these data sets. The top panel shows the distribution of $R$ values attained from these data simulations. The confidence in the model selection at any particular $\mathcal{R}$ (bottom panel) is determined by the relative heights of the two distributions at that $\mathcal{R}$ . The current data strongly favor the KCTF model. . . . .	31

4.3	The posterior distribution of $f$ for each of our two-migration-channel models, conditioned on 32 observed $\lambda$ values. We show that if the true misaligned population were misaligned according to the KCTF distribution, then we expect that about 64% of close-in planets are misaligned. Similarly, if the misaligning mechanism were PSTF, then the most likely value of $f$ is close to 55%. Strictly speaking, this figure shows the likelihood of the current data as a function of $f$ for each model; if these curves were normalized they would be proper probability distributions, given our flat prior for $f$ . However, plotting them in un-normalized form is illustrative, since we use the ratio of the areas under these curves (the marginal likelihood) as our two-mode model selection criterion. . . . .	33
4.4	The relationship between our model selection statistic and model selection confidence for comparing our two-mode migration models, where only a fraction $f$ of planets are misaligned and the rest are in perfectly-aligned systems. $\mathcal{R}$ in this case is the logarithm of the ratio of the marginal likelihoods of the two models, and the $\mathcal{R}$ distributions are generated by measuring the $\mathcal{R}$ values for data simulations at different $f$ . The PSTF + aligned model is preferred over the KCTF + aligned model. See §4.4.2 for details. . . . .	34
4.5	The relationship between our model selection statistic and model selection confidence, comparing the preferred single-mode migration scenario (KCTF) to the preferred two-mode model (PSTF + aligned). PSTF + aligned is strongly preferred over single-mode KCTF. . . . .	35
4.6	The relationship between our model selection statistic and model selection confidence, but using only the 12 hot stars ( $T_{\text{eff}} > 6250$ ) as our data. One-channel migration favors the PSTF mechanism for this subset of the data, but not conclusively. The $\mathcal{R}$ thresholds required to reach 95% confidence in either model for this sample size are marked. . . . .	37
4.7	The posterior distribution of $f$ for each of our two-migration-channel models, conditioned on the 12 observed hot star ( $T_{\text{eff}} > 6250$ ) $\lambda$ values. As Winn et al. (2010) point out, hot stars appear to be misaligned more often than the overall stellar population. Because $f = 1$ has a high probability under both models, we do not further pursue two-mode migration model comparison for this subsample. . . . .	38
4.8	The probability with which a given-sized future sample of hot-star ( $T_{\text{eff}} > 6250$ K) $\lambda$ values (in addition to the 12 that currently exist) will result in a confident selection ( $> 95\%$ ) of one of our misalignment models over another. We use data simulations of various sample sizes (described in §4.6) to define the confidence thresholds and to determine how likely those thresholds are to be reached for each particular sample size. For example, with 28 more measurements (40 total), we have a 75% chance of selecting the PSTF model at $> 95\%$ confidence. . . . .	39

5.1	The probability for a possibly blending star to be projected within $2''$ of a <i>Kepler</i> target star, as a function of Galactic latitude, as determined by TRILEGAL simulations. The plotted points are simulations; the lines are the exponential fits as described in Equation 5.8. . . . .	47
5.2	The probability for a possibly blending star to be projected within $2''$ of a <i>Kepler</i> target star, as a function of both Galactic latitude and target star magnitude, as determined by TRILEGAL simulations. . . . .	48
5.3	The false positive probability of a <i>Kepler</i> candidate, according to our basic framework (i.e. independent of $\delta$ ), as a function of target star magnitude $m_K$ and galactic latitude. A planet occurrence rate of 20% is assumed. This plot assumes that <i>Kepler</i> is able to internally restrict the radius inside which a possible blended binary might reside to $2''$ . There is a small gradient across the field, but the false positive probability is uniformly low. . . . .	52
5.4	False positive probability as a function of assumed planet occurrence rate, for a $m_K = 14$ target star in the center of the <i>Kepler</i> field. The occurrence rate of planets detectable by <i>Kepler</i> is not known for sure, but RV surveys, especially the <i>NASA-UC Eta-Earth Survey</i> of Howard et al. (2010), have made inroads in measuring the fraction of stars hosting low-mass planets. The hashed area below 9% represents the occurrence rate of planets with $P < 50$ days that is ruled out with 95% confidence by $\eta_{\text{earth}}$ , counting only the firm detections, and not correcting for completeness. The central hashed area represents the 95% confidence region calculated, including candidate planets and completeness correction, for minimum masses greater than $3 M_{\oplus}$ . Extrapolating their observed mass distribution down to $0.5 M_{\oplus}$ brings their total estimated planet occurrence rate to 43%. Overall, this plot shows that our derived FPP cannot reasonably be any higher than 3% if our planet occurrence estimate is incorrect, and will likely be lower. .	54
5.5	Distributions of apparent “transit” depths $\delta$ for different scenarios. The blended binary and hierarchical triple distributions are based on TRILEGAL simulations with the binary distribution assumptions discussed in §5.2. Examples of $\delta$ distributions are given for different target star properties, showing how the blended binary scenario depends on target star apparent magnitude and how the hierarchical triple distribution depends on intrinsic target star mass. The planet distribution comes from an assumption of a continuous power law in planet radius $dN/dR_p \propto R_p^{-2}$ , including random statistical dilution by binary companions. Note how blended binaries become less significant for deep signals and how eclipsing triples become insignificant for the shallowest signals. The deep end of the false positive $\delta$ distributions are mostly due to eccentric binaries that are oriented such that only a single eclipse occurs. . . . .	56

- 5.6 As stars get fainter and transit signals get shallower, the ability for *Kepler* to observe a centroid shift indicative of a displaced blended eclipsing binary decreases. We parametrize this effect according to Equation 5.19. The plateau towards shallow depths is a result of the maximum blending area for this example being set to an aperture of 8 *Kepler* pixels; the location of this plateau for any particular target will depend on its aperture size. This plot is made according to a Galactic latitude in the middle of the *Kepler* field; other latitudes will scale appropriately according to the varying stellar density. The planet radii are marked assuming a Solar-radius star. 57
- 5.7 The probability that a 10-day signal of a given depth will be a false positive, shown for both an  $m_K = 11$  and an  $m_K = 15$  star with Solar properties, Galactic latitude in the center of the *Kepler* field, and an 8 pixel *Kepler* aperture. A overall planet occurrence rate of 20% and a planet radius function  $dN/dR \propto R^{-2}$  are assumed. The false positive probability increases towards deeper signals mostly because of the decreasing planet radius function combined with the significant tails of the false positive  $\delta$  distributions. For the fainter star the false positive probability begins to pick up again at the shallowest depths as it becomes more difficult for *Kepler* to rule out displaced blended binaries via astrometry. The plateau corresponds to the point at which the blending radius becomes equal to the aperture size. The dotted line represents the effect of restricting the blending radius to  $2''$  with a single high-resolution image; this can decrease the false positive probability for Earth-sized signals from  $\sim 4\%$  to  $< 1\%$ . . . . . 58
- 5.8 These plots illustrate the behavior of *Kepler* false positive probability (FPP) as a function of target star magnitude ( $m_K$ ) and Galactic latitude for three particular choices of transit depth  $\delta$ , all plotted with the same color scale. A planet occurrence rate of 20% is assumed, and the target star is fixed to have Solar mass, Solar radius, and a photometric aperture of 8 pixels. These plots are similar to Figure 5.3, except for that they take into account both the changing blend radius as a function of  $m_K$  and  $\delta$  (Equation 5.19) and the relative likelihoods of false positives and planets at the chosen values of  $\delta$ . All three  $\delta$  values show increasing FPP towards fainter target stars and lower galactic latitudes, though the strength of the gradient decreases for the deeper signals, as the relative importance of the hierarchical triple scenario increases. Dotted lines show the FPP contours if the blend radius were restricted to  $2''$ , illustrating the power of a single deep high-resolution observation. Note that only in the Earth-sized transit case does the high-resolution observation result in  $FPP < 0.01$ , as the FPP for shallower signals is dominated by chance-alignment blended binaries while the FPP for deeper signals is dominated by hierarchical triple scenarios. . . . . 59

5.9	The distribution of false positive probabilities (FPPs) among the 1235 <i>Kepler</i> planet candidates announced in Borucki et al. (2011). FPP for each candidate is calculated individually, taking into account the apparent Kepler magnitude, Galactic latitude, mass and radius of the host star, the depth and period of the signal, the number of pixels contained the optimal aperture used for <i>Kepler</i> photometry, and the detection threshold of each KOI time series. Nearly 90% (1098) have FPPs less than 10%, and over half (713) have FPPs less than 5%. The mean FPP of the sample is about 6%, indicating that we expect there to be fewer than $\sim 75$ false positives among the candidate sample. An important caveat here is that these calculations assume that all candidates have passed the false positive-vetting tests that are possible using <i>Kepler</i> photometry.	60
6.1	The distribution of hierarchical triple systems allowed for Kepler-19b (KOI 84.01) by the KIC $g-r$ and 2MASS $J-K$ colors. Note how there is a tail of higher-mass systems that can mimic Solar-like colors. This is important for calculating FPP, because systems with higher-mass (and thus higher-luminosity) primaries are more prone to cause false positives with shapes similar to planet transits. If the spectroscopic properties of the primary are measured (e.g. with a high-resolution spectrum), then this significantly constrains the false positive landscape (see Figure 6.6).	72
6.2	The specific occurrence rate $f_p$ as a function of radius for three examples of generic radius distributions, with all scaling as the inverse square of the planet radius. I define the specific occurrence rate at a particular radius $R_p$ to be the integral of the planet radius probability density function between $0.7R_p$ and $1.3R_p$ (Equation 6.4). Even though the use of the specific occurrence rate in the analysis presented in this paper makes it independent of assumptions of the overall planet rate or the exact shape of the radius distribution, this illustrates what reasonable values of $f_p$ might be. For example, it is clear that $f_p = 0.01$ is quite conservative for nearly any planet size, especially for small planets.	76
6.3	The phase-folded <i>Kepler</i> photometry of Kepler-19b (KOI 84.01). The solid line illustrates the best-fitting trapezoid model used for the likelihood calculations (§6.3.3). Note that the transit is clearly box-shaped, not V-shaped.	77

- 6.4 The false positive landscape for Kepler-19b (KOI 84.01). Each of these plots illustrates a three-dimensional probability distribution for the trapezoidal shape parameters (depth  $\delta$ , duration  $T$ , and “slope”  $T/\tau$ ) for a false positive scenario (see §6.2 for descriptions of the scenarios). Each of these distributions is made by simulating a statistically representative population (§6.3.1) for a scenario and fitting the shape parameters to each simulated instance of the scenario. Each population begins with 20,000 simulated instances, and only instances that pass all available observational constraints are included in these distributions—in this case the constraints are that the blended star be at least 1 mag fainter than the primary target, the lack of an observed secondary eclipse deeper than 200ppm, and that the KIC  $g-r$  and 2MASS  $J-K$  colors match within 0.1 mag. The pie chart for each scenario illustrates what fraction of the initial simulations pass these tests. On each plot the shape parameters of the transit signal are marked, with the ‘X’ showing the median of an MCMC fit, and error bars illustrating 95% confidence. The probability of each scenario is calculated by dividing the prior  $\times$  likelihood for that particular scenario by the sum of prior  $\times$  likelihood for all the scenarios (including the planet scenario illustrated in Figure 6.5). Priors for each scenario are calculated according to §6.3.2, and the likelihoods are calculated by integrating the illustrated distributions over the observed measurement. . . . . 85
- 6.5 The transiting planet scenario for Kepler-19b (KOI 84.01). The plots illustrate the three-dimensional probability distribution for the trapezoidal shape parameters for simulations of a representative population (§6.3.1) of transiting planets around KOI 84. The radius distribution of planets in this population is chosen to be a narrow range centered on the radius derived for KOI 84.01, assuming it is a *bona fide* transiting planet. The spread and shape of this distribution are caused by different-sized planets in the simulation, different stellar properties (according to uncertainties in stellar radius and mass), and by variations in orbital inclination, which changes duration and slope in a correlated manner (bottom-right panel). The assumed occurrence rate of planets in this radius bin for this calculation is 1% (see §6.3.4 for more discussion about this “specific occurrence rate”). The probability of the Kepler-19b signal being a transiting planet, given the false positive landscape illustrated in Figure 6.4 (constrained by only the signal photometry and KIC colors) is about 97%, giving FPP = 0.03. The signal does not fall quite exactly in the middle of the  $T/\tau$  distribution because this population simulation uses photometrically estimated physical properties (mass, radius) for the the host star, rather than the more accurate spectroscopically derived properties. Note that uncertainties like this will tend to decrease the planet likelihood, and thus will contribute to *overestimating* the FPP. 86



- 6.6 The hierarchical triple eclipsing binary (HEB) scenario for Kepler-19b (KOI 84.01) under the constraint of spectroscopic characterization of the host star. The plots illustrate the three-dimensional probability distribution for the trapezoidal shape parameters for simulations of a representative population (§3.1) of HEB false positive scenarios, but only including those instances in which the spectroscopic properties ( $T_{\text{eff}}$ ,  $\log g$ ) of the primary star match those that were measured for KOI 84. Compare this distribution to Figure 6.4(c): the systems whose shape parameters overlapped with the properties of the measured signal are not allowed by the spectroscopic characterization—these correspond to the higher-mass systems illustrated in Figure 6.1. This constraint allows for the signal to be validated, as the HEB scenario was the most likely FP scenario before including the follow-up spectral information. . . . . 87
- 6.7 The phase-folded *Kepler* photometry of KOI 552.01. The solid line illustrates the best-fitting trapezoid model used for the likelihood calculations (§6.3.3). Bouchy et al. (2011) measured large RV variations of the host star, suggesting that this signal is caused by an eclipsing M-dwarf rather than a planet. Note that the signal is clearly more V-shaped than box-shaped. . . 87
- 6.8 The hierarchical triple eclipsing binary system scenario for KOI 552.01. This KOI was measured by Bouchy et al. (2011) to have  $\sim$ km/s radial velocity variation between two epochs, suggesting a stellar eclipse rather than a transiting planet—an observed astrophysical false positive. The analysis presented in this paper gives an FPP of  $>90\%$  for this signal, and this figure illustrates why: the shape of the signal is exactly consistent with what would be expected from a hierarchical triple system including an eclipsing binary. . . . . 88
- 6.9 The results of testing the effects of sample variance and changing various assumptions on the false positive calculations described in this paper. Each colored vertical line represents the full range of  $f_{p,V}$  (the specific occurrence rate assumption required for validation) over three iterations of the procedure for each planet from Table 6.1, ordered by increasing radius. The five different vertical lines for each planet represent five different experiments, varying the size of the Monte Carlo simulations and different input astrophysical assumptions. The horizontal black lines represent the value of  $f_{p,V}$  inferred from an assumption of an overall 40% planet occurrence rate with  $dN/dR \propto R^{-2}$ . Below the line means the signal is securely validated. All of the planets that were claimed as validated by the original analysis (the final column of Table 6.1) remain validated for each iteration of each of these tests. . . . . 88

- 7.1 Evidence supporting the hypothesis that small planets are incomplete in the Cool KOI sample. The solid black line is the observed (smoothed) distribution of planets smaller than  $1 R_{\oplus}$ ; the grey shaded area is the observed period distribution of all the Cool KOIs. Neither distribution is corrected for transit probability. The vertical dashed red line indicates the period at which a  $1 R_{\oplus}$  planet around a  $0.5 R_{\odot}$  star (typical of the Cool KOI sample) would have SNR of 7.1, the nominal detection threshold for KOI identification. The lack of observed small planets at periods longer than 10 days is thus very plausibly due to incompleteness. . . . . 90
- 7.2 The period distribution of planets around *Kepler*'s M dwarfs. The grey shaded region is the implied period distribution of all planets combined, correcting for the effects of transit probability. The bar charts show the observed numbers of planets of different sizes in each period bin. Note the declining fraction of small planets as a function of period—this is most likely an effect of declining detection efficiency for smaller planets on longer-period orbits, and this must be properly accounted for when constructing the planet radius function. The radius function calculation in this paper assumes that all planets are distributed according to the shaded distribution, regardless of planet radius. See §7.5 for a discussion of this assumption. . . . . 96
- 7.3 Two examples of the SNR distributions resulting from simulating the transit of a given planet around every target star, given randomly assigned periods and impact parameters. The properties of this distribution depend on the properties of the detected system, and the integral of the pipeline detection efficiency function over this distribution gives the “discovery fraction”  $\eta_{\text{disc}}$ . KOI-961.01, a sub-Earth-sized planet in a very short orbit around a very small star (Muirhead et al. 2012b), would have been detectable in only about 1/6 of potential configurations, whereas KOI-952.03, a larger planet around a larger star (Swift et al. 2013), would be detectable in almost any configuration, even though its actual SNR is smaller than that of KOI-961.01. . . . . 97

- 7.4 The empirical radius distribution of planets orbiting M dwarfs with periods  $<90$  days (black continuous curve), estimated with a weighted kernel density estimator (wKDE; see §7.3.3), with the bootstrap resampling-derived  $1\sigma$  uncertainty swath shaded grey—essentially a running Poisson error bar. The detection efficiency as a function of signal-to-noise ratio has been quantified by an SNR ramp from 0 at SNR = 6 to 1 at SNR = 12. The blue horizontal lines represent the standard “occurrence rate per bin” calculations for this sample. The vertical red lines represent the radii of individual planets in the sample, with their heights in proportion to the weight factors  $w_i$ . The green dotted curve is an  $R^{-2}$  power law, which bears a striking (and uncontrived) resemblance to the shape of this non-parametric radius function between about  $1.25$  and  $2 R_{\oplus}$ ; below about  $1.25 R_{\oplus}$ , the distribution appears to level off, and turn over below  $1 R_{\oplus}$ . Between about  $2$  and  $2.5 R_{\oplus}$  there appears to be an excess over a smooth distribution; this may be caused by a significant population of planets with H/He atmospheres. There is an average of 1.5 planets per cool star in orbits  $<90$  days over this radius range, and there is an average of greater than 0.5 planets per cool star in this period range for radii between  $1$  and  $1.5 R_{\oplus}$ . . . . . 98
- 7.5 The planet radius distribution for  $P < 50$  days, using the Batalha et al. (2012) catalog in order to compare to previous studies, demonstrating the effect of the corrections accounted for in this work. The continuous curves are the non-parametric empirical density estimates, and the horizontal blue lines are the occurrence rates per bin, using the same bins as Dressing & Charbonneau (2013). The vertical red lines represent the radii of individual planets in the sample, with their heights in proportion to the weight factors  $w_i$ . The non-solid linestyles represent different analysis methods. Whereas Method 1 (solid lines) uses the full analysis described in this paper, with detection efficiency described by an SNR ramp (Fressin et al. 2013) and  $\phi_{\text{SNR},i}$  constructed by assigning random periods (§7.3.2), Methods 2 (dashed) and 3 (dotted) both keep period fixed when constructing  $\phi_{\text{SNR},i}$ , and Method 3 uses an SNR = 7.1 detection threshold rather than the SNR ramp. Method 2 is similar to the occurrence calculation in Fressin et al. (2013) and Method 3 uses the methods employed in Dressing & Charbonneau (2013). The importance of both a well-characterized detection efficiency function and treating period-based incompleteness correctly is clear: incorporating both these considerations significantly changes both the qualitative shape (especially as visualized with histograms) and normalization of the radius function below  $2 R_{\oplus}$ . In particular, the Dressing & Charbonneau (2013) analysis underestimates the occurrence rates of planets between  $0.5$  and  $1.4 R_{\oplus}$  by about a factor of two. 100

8.1	The radius and period distributions of planets around solar-type stars (4100-6100 K), derived from the distribution of <i>Kepler</i> planet candidates by Howard et al. (2012). These results depended on assuming that all the candidates were planets; i.e. that the false positive rate was negligible. . . . .	105
8.2	Summary of the derivation of the underlying shape of the exoplanet joint period-radius distribution, assuming a joint power-law distribution $\propto R^\alpha P^\beta$ (Youdin 2011). The population of planet candidates from Borucki et al. (2011) is split into four quadrants, short/long-period (split at 7 days) and small/large (split at $3 R_\oplus$ ). The most striking feature found here is the qualitatively different shape of the radius function for short and long period planets. In addition (not indicated in these plots), extrapolation of these power-law fits indicates that there could be an average of $\approx 3$ Earth-sized planets per star out to 365-day orbits. . . . .	106

# List of Tables

4.1	Results summary . . . . .	40
5.1	Polynomial coefficients <sup>1</sup> for Equation 5.15 . . . . .	51
6.1	Test sample: Non-TTV Confirmed Kepler planets <sup>1</sup> . . . . .	79
6.2	Test sample: False positives <sup>1</sup> . . . . .	81
7.1	Data Used in Radius wKDE . . . . .	101

## **Part I**

# **Introduction**

Extrasolar planetary systems are very common—there are at least several times more planets than stars in the Galaxy. Small planets are much more common than large ones. Many stars host multiple planets on short orbits—as many as four or five within the orbit of Mercury. A large number of these planets are unlike anything in our Solar System—between the size of Earth and Neptune. Planets may even commonly form around binary stars. Stars less than half a Solar mass often host planets the size of Earth and even smaller, and sometimes multiple such planets orbit within just a few hundredths of an Astronomical Unit (AU). All of these demographic facts about exoplanets have been discovered in just the last three years, based on data from NASA’s *Kepler* mission. The research that I present in this dissertation includes key results that helped make these conclusions possible, as well as analysis that lays the groundwork for future discoveries of exoplanet demographics.

This introductory part contains three chapters: Chapter 1 gives an overview of the landscape of exoplanet demographics as it existed before *Kepler*. Chapter 2 gives a background about transiting planets and introduces the *Kepler* mission. Chapter 3 gives a brief overview of Bayesian inference, a statistical methodology which I draw upon over the course of the dissertation.

## Chapter 1

# The History and Goals of Exoplanet Demographics

The scientific goal motivating the work presented in this dissertation is to understand the demographics of exoplanet systems. What this means, most precisely, is characterizing as fully as possible the multi-dimensional probability distribution function for planets:

$$f_p(M_p, R_p, P, e, i, M_*, [\text{Fe}/\text{H}], N_*, N_p \dots). \quad (1.1)$$

In other words, how are planets distributed as a function of their physical ( $M_p$ ,  $R_p$ ) and orbital ( $P$ ,  $e$ , and inclination  $i$ ) properties, and of the properties of their host stars (e.g.,  $M_*$ ,  $[\text{Fe}/\text{H}]$ ), and of their environmental properties (e.g., host star multiplicity  $N_*$ , or number of other planets in the system  $N_p$ )? The ultimate goal of this effort is to use the ensemble of exoplanet observations to meaningfully inform theories of planetary system formation and evolution in the hope of understanding the origin of the diversity of planetary systems. Along the road to this goal also lie answers to deep questions about how “typical” our Solar System is, and how common potentially habitable planets might be in the Galaxy.

The first giant exoplanets to be discovered taught the first important lesson about their distribution function: giant planets can exist in very short orbits around their stars (so-called “hot Jupiters”). The implication of this discovery is that the orbits of at least some planets must evolve considerably from where they form, as there is no known mechanism to form giant planets so close to their host stars. This invited several different explanations for planet “migration,” and this remains an active area of research, recently reinvigorated by the more recent discovery that a substantial fraction of these hot Jupiters appear to orbit their stars on highly inclined trajectories (Hébrard et al. (2008); Winn et al. (2009a,b); Johnson et al. (2009); Narita et al. (2009); Triaud et al. (2010), and see Chapter 4).

As many more giant planet detections followed the surprise of these first unexpected ones, the ability and desire to learn more about their demographics increased. This endeavor may be thought of at different levels of sophistication. The most shallow (though necessary) level might be considered simply bookkeep-



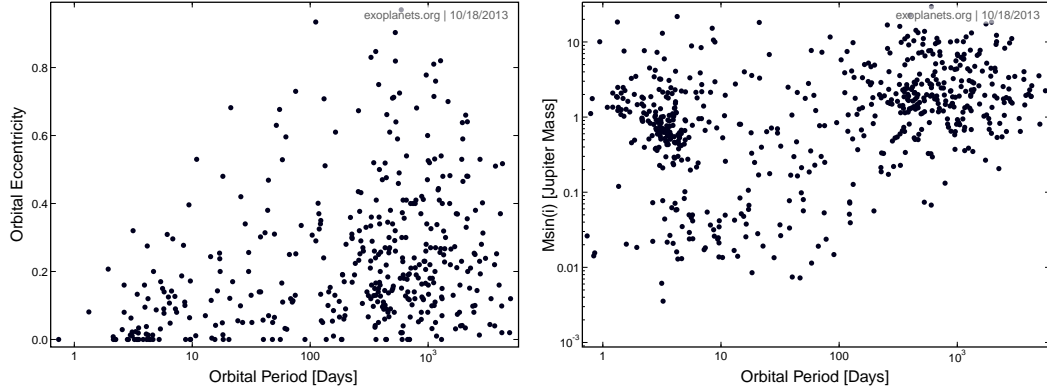


Figure 1.1 Unfiltered distributions of measured exoplanet quantities from the Exoplanet Orbit Database at [exoplanets.org](http://exoplanets.org), demonstrating both the power and the danger of inspecting simple scatter plots in the service of exoplanet demographics. The left-hand panel shows very clearly that many exoplanets likely have more violent dynamical histories than the circular planets of our solar system. The right-hand panel shows the minimum masses ( $M \sin(i)$ ) and periods of all known planets; however, proper interpretation of this plot is more difficult because of the various different detection biases at work. Larger and shorter-period planets are easier to detect, and especially transit surveys (which are represented here, in addition to radial velocity surveys) are very heavily biased toward detecting large and close-in planets. However, the increase in the density of detected giant planets beyond orbital periods of  $>200$  days is a true feature of this distribution.

ing, as much of the history of astronomy has been about carefully observing and categorizing the Universe. While initial qualitative distinctions may be made with little or no thought as to physical interpretation, often theoretical understanding follows, sometimes at a very deep level. A good example of this is the Hertzsprung-Russell (H-R) diagram, which was first a purely observational construction relating stars' colors to their absolute magnitudes, but is now known to trace at a deep level how stars work and how they evolve. In other words, simply making lots of careful measurements of exoplanets and plotting various measured quantities against each other has the potential to reveal important insights.

To illustrate this, Figure 1 shows two examples of exoplanet quantities being plotted against each other, using the data and tools available at [exoplanets.org](http://exoplanets.org) (Wright et al. 2011). In each of these panels is plotted all exoplanets with well-measured orbits, meaning in this case planets whose RV signal has been detected (whether originally detected in RV or transit). The juxtaposition of these two scatter plots demonstrates that sometimes physical interpretation is evident from an exercise like this, whereas sometimes it is not. The left-hand panel plots orbital period against orbital eccentricity. If all one knew about planetary systems was our solar system with its neatly ordered, nearly circular orbits, this would be a surprising plot indeed. What it indicates very clearly is that many planets in other planetary systems very likely have had much different dynamical histories than our Solar System. This has been explored theoretically in detail, with many different studies showing that chaotic gravitational interactions between multiple giant planets in the same system can qualitatively reproduce the observed eccentricity distribution (Jurić & Tremaine 2008; Ford & Rasio 2008; Chatterjee et al. 2008). This is an example of a simple scatter plot of measured quantities that gives an important qualitative physical insight into the nature of planetary systems.

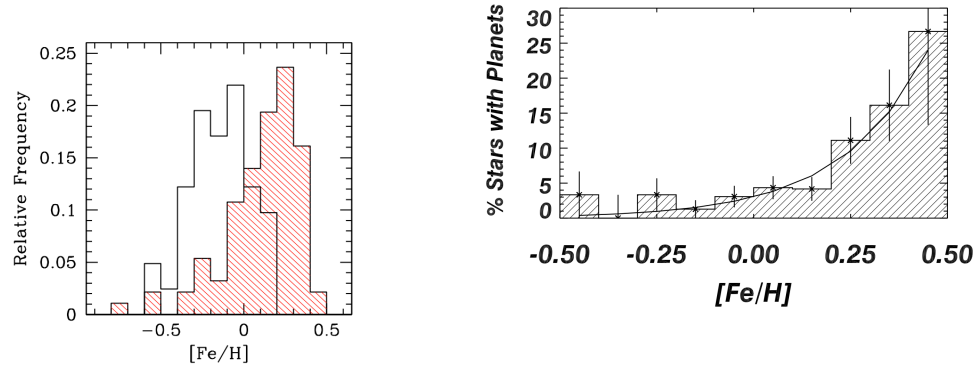


Figure 1.2 Two different views of the planet-metallicity relation. The left panel (Santos et al. 2004) shows the distribution of the metallicity of stars with known planets c. 2004 (red-hashed histogram) compared to a volume-limited comparison sample of stars (open histogram), indicating that planet-hosting stars are preferentially metal-rich. The right-hand panel (Fischer & Valenti 2005) quantifies this same relationship more precisely, measuring the occurrence rate (number of detections divided by total number of stars surveyed) in different metallicity bins.

Inspection of the right-hand panel indicates several things: there are many more massive planets than small planets, smaller planets only exist at short periods, and giant planets have a strongly bimodal period distribution. All three of these conclusions are incorrect, however, and understanding where the data come from is thus crucial to proper interpretation. Dominating the structure of this plot, of course, are detection biases of several sorts. First of all, large and short-period planets are much more easily detected than smaller and long-period ones, meaning that the decrease in density of points to the bottom and right of this plot is not reflective of reality. In addition, this plot includes planets detected by both RV and transit surveys, and the detection efficiency of transit searches is a much stronger function of planet size and period than RV surveys, which is the dominant reason for the large clump in the upper left and the apparent period bimodality. On the other hand, the paucity of giant planets with orbits between 10 and 100 days (compared to those with  $\sim 200$ -1000 day orbits) is indeed significant, because shorter-period planets are more detectable. This likely does point to something physical about the formation and orbital evolution of giant planets. This exercise demonstrates that this first simple level of demographic analysis, though occasionally revealing, does not work for long before a more sophisticated approach is necessary.

An example of such an approach is analysis of *a single survey*—including detections and non-detections—to draw conclusions about exoplanet demographics, rather than attempt to do so based simply on a list of detections from heterogeneous surveys. This is especially important when attempting to answer questions like how the occurrence rate of planets might depend on stellar properties. A classic example of this is the now well-known planet-metallicity correlation (Figure 1). While a simple histogram of the metallicity of detected planets (LH panel) does hint that planets prefer metal-rich stars (as the distribution of stellar metallicities in the stellar neighborhood is considerably more metal-poor than the distribution of planet hosts),

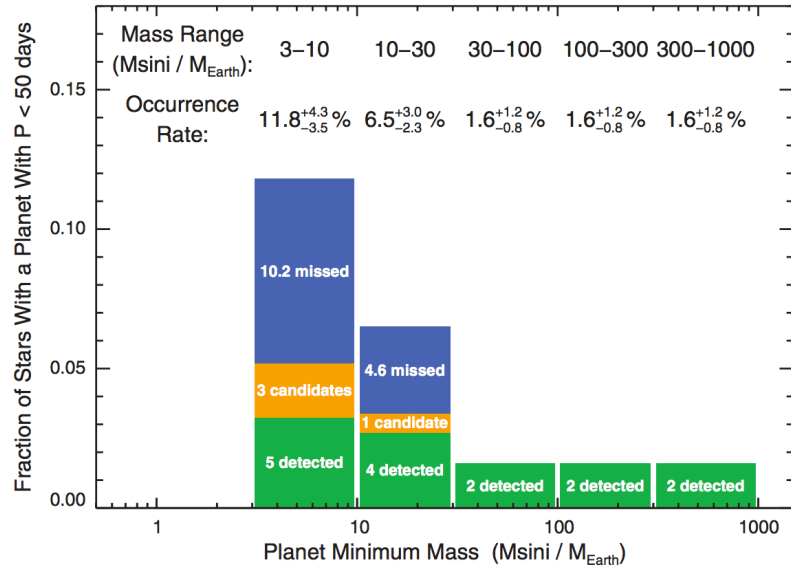


Figure 1.3 Results of the *NASA-UC eta-Earth* survey (Howard et al. 2010), which was a high-precision RV survey of 166 FGK dwarf stars intended to measure the occurrence rate of planets down to much lower masses than had previously been done, showing that “super-Earth” planets are much more common than giant planets. The key to accomplishing this was to precisely quantify the fractional completeness of the survey, enabling correction for “missed planets.”

this does not come close to revealing what is actually happening. The RH panel, on the other hand, based on analysis of the detections *and non-detections* of a single survey, successfully distills the essence of the issue: the probability for a star to host a giant planet is a demonstrably strong function of metallicity. This is exactly as predicted by the “core accretion” theory of giant planet formation (Mizuno 1980; Pollack et al. 1996), which argues for bottom-up formation of Jupiter-like planets, rather than “top down” direct collapse via gravitational instability (Boss 1997).

Another step forward is to consider not only detections and non-detections in a survey, but to carefully quantify *fractional* completeness. A good example of this is the *NASA-UC eta-Earth* survey (Howard et al. 2010), the goal of which was to characterize the occurrence rate of planets down to much lower masses than had previously been accomplished. The survey was carefully designed and executed, making precise Doppler measurements ( $\sim 1$  m/s precision) with Keck/HIRES of 166 G- and K-type stars, ensuring that each target got at least 20 RV observations over five years, including at least one group of 6 to 12 measurements within a 12-night span to maximize sensitivity to short-period planets. This uniformity enabled the authors to determine completeness  $C$  at each point in the mass-period plane, such that a planet detected in a region with  $C < 1$  would count as  $1/C$  planets in the final tally. Figure 1 illustrates the results of this survey, which found a steeply rising planet mass function below 30 Earth masses—about 20% of Sun-like stars host planets

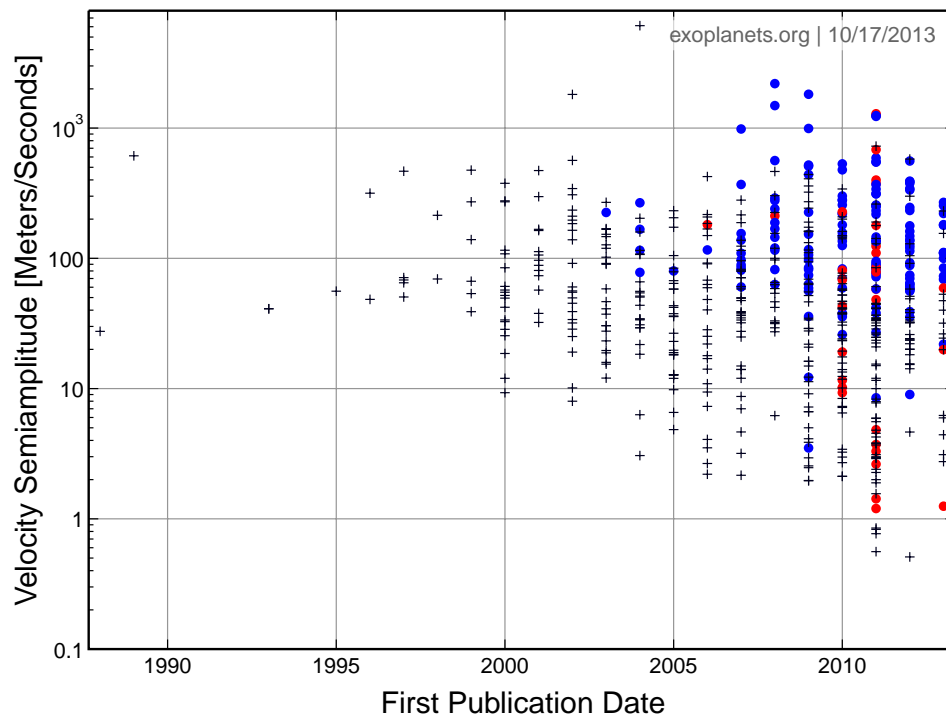


Figure 1.4 The history of exoplanet discovery. Black crosses indicate planets detected in RV surveys, blue dots indicate planets detected in transit surveys, and red dots indicate planets detected by the *Kepler* mission. Despite the evident advances in RV instrumentation, the vast majority of planets with measured velocity amplitudes remain giant planets (Jupiter’s perturbation on the Sun has a semi-amplitude of about 10 m/s).

between 3 and 30 Earth masses with periods less than 50 days, compared to a  $\sim 5\%$  occurrence rate for planets larger than 30 Earth masses (about 0.1 Jupiter masses).

This finding that small, “super-Earth”-type planets are significantly more common than giant planets is not surprising, but it underscores the fact that, before the *Kepler* mission (which I will describe in detail in Chapter 2), all efforts toward understanding exoplanet demographics had just explored only the very tip of the iceberg. A striking way to quantify this is to note that in the *exoplanets.org* database there are 500 planets with minimum mass  $>100$  Earth masses ( $0.3 M_{\text{Jup}}$ ), and only 36 with measured minimum mass between 3 and  $10 M_{\oplus}$ . On the other hand, the *eta-Earth* survey found that planets between 3 and 10 Earth masses are actually about  $4\times$  more common than planets  $>0.3$  Jupiter masses. If “super-Earths” were detected at the same level of completeness as the giant planets, there should be  $\sim 2000$  of them in order to complement the 500 giants. In other words, the 3-10  $M_{\oplus}$  range has been sampled at only about 2% of the completeness of the  $>0.3 M_{\text{Jup}}$  range.

This incompleteness at low masses can be seen in the distribution of points in Figure 1, and the reason for it can be seen in the exoplanet discovery history shown in Figure 1. For the first decade after the first confirmed announcement of an exoplanet around a main sequence star (Mayor & Queloz 1995), nearly all the detected exoplanets had RV semi-amplitudes considerably larger than 10 m/s (approximately the semi-

amplitude with which Jupiter perturbs our Sun). While the lower envelope of detectability has continued to drop (the semi-amplitude of a  $10 M_{\oplus}$  planet in a 50-day orbit around a Sun-like star is only 1.7 m/s), the vast majority of well-characterized exoplanets are still giants.

To conclude this chapter, I summarize the key demographic findings of the RV survey-dominated era of exoplanets:

- Giant planets exist where they should not be able to form. These “hot Jupiters” occur around about 1% of Sun-like (FGK) stars (Cumming et al. 2008; Wright et al. 2012).
- About 10% of FGK dwarfs host a giant planet  $M > 0.3M_{\text{Jup}}$  and  $P < 2000\text{d}$ . (Cumming et al. 2008).
- The probability that a star hosts a giant planet is a strong function of metallicity (Santos et al. 2004; Fischer & Valenti 2005), and also an increasing function of stellar mass (Johnson et al. 2010a).
- Giant planets display a wide range of eccentricities, likely indicating that their orbits have been shaped by multi-body interactions (Jurić & Tremaine 2008; Chatterjee et al. 2008; Ford & Rasio 2008).
- The orbits of many hot Jupiters are misaligned with their stellar spin axes (Triaud et al. 2010).
- The occurrence rate of planets increases toward smaller planet mass (Howard et al. 2010).
- Multi-planet systems occur, but are not a majority of giant planet systems: about 28% of stars with RV-detected planets show significant evidence of a second companion (Wright 2010).

In the next chapter I will discuss how the *Kepler* mission has enabled a revolution in this landscape of exoplanet demographics.

## Chapter 2

# Transiting Planets and the *Kepler* Revolution

When an exoplanet's orbit is nearly perpendicular to the plane of the sky, it will eclipse its host star and occult a small fraction of the star's light. This is called a *transit*, and a planet in this configuration reveals two otherwise inaccessible quantities: its radius (from the depth of the transit) and the orientation of its orbit. These quantities are most powerful when combined with a radial velocity (RV) detection, which typically can only measure a planet's minimum mass ( $M \sin i$ , where  $i$  is the planet's orbital inclination), but for a transiting planet reveals its true mass, since the inclination is known. The mass and radius together give a planet's density, which in turn constrains its composition. There is also a host of additional more detailed investigations that can be pursued when a planet is known to transit, such as studies of atmospheric composition and circulation, or measuring the Rossiter-McLaughlin (R-M) effect, which can reveal the relative orientation of the planet's orbit and the host star's spin axis (see Chapter 4 for more detail).

While the fortuitous orientation of a transiting planet makes it amenable to detailed study, the fact remains that a transiting planet requires fortuitous orientation, and so only a small minority of exoplanets transit. The probability for a planet's orbit to be aligned such that its transit is equal (for a circular orbit) to the ratio of the host star's radius to the planet's orbital distance: for a Jupiter-sized planet in a 3-day orbit around a Sun-like star this probability is about 1 in 10; for an Earth-sized planet in a year-long orbit, 1 in 200. Consequently, finding transiting planets requires monitoring a large number of stars to have the chance of detecting any.

The photometric signal of a transiting planet is also quite challenging to detect. A star dims by a fractional amount approximately equal to the ratio of the area of the planet's disk to the area of the stellar disk—for a Sun-like star this fractional depth is about 1% for a Jupiter-sized planet, 0.1% for a Neptune-sized planet, and 0.01% for an Earth-sized planet. A transit survey therefore must not only monitor a large number of stars but do so at a level of photometric precision sufficient to detect the desired planets.

The biggest challenge of a transit survey, however, is neither the number of stars it must monitor nor the photometric precision it must attain, which are relatively easily overcome with wide-field cameras and modern CCD technology and data reduction algorithms. Rather, the biggest challenge transit surveys face

is that finding periodic stellar dimmings from wide-field precise photometry is only the first step in a long process to securely identify a transiting planet. The main reason for the necessity of this long process is the preponderance of astrophysical false positives.

## 2.1 False Positives in Transit Surveys

Strictly speaking, the signature of a transiting planet is a small periodic decrease in the flux of photons impinging on the pixels of a CCD defined to be the *photometric aperture* of a given target star. An astrophysical false positive is any of a number of eclipse scenarios that cause such a signal without being the hoped-for planet. This may be a dwarf star eclipsing a giant, a grazing eclipsing binary, or a so-called *blended eclipsing binary*, all of which can cause dimmings similar to that of a transiting planet. A blended binary scenario occurs when a faint eclipsing binary system lies somewhere within a target star’s photometric aperture. For example, there might be an eclipsing binary pair where the secondary is about one-third the radius of the primary, thus causing a fractional eclipse depth of about 10%. If this system happens to land on a pixel within the aperture of a star  $10\times$  brighter, then the photometric time series of the brighter star will appear to dim periodically by only 1%—similar to a Jupiter-sized planet transiting a Sun-like star. This phenomenon, where the presence of unvarying “third light” changes the apparent fractional amplitude of a photometric signal, is known as *dilution*.

The classic example of an astrophysical false positive is the case of OGLE-TR-33 (Torres et al. 2004). Identified by the Optical Gravitational Lensing Experiment (OGLE) as a transiting planet candidate with a periodic photometric dip of 2%, subsequent follow-up observations seemed only to confirm its planetary nature: periodic RV variations were discovered in phase with the transit signal, no nearby stars were observed to be blended within the photometric aperture, and even the eclipse depth was measured to be color-independent. However, further investigation revealed that various blended eclipsing binary scenarios could fit the light curve just as well as a transiting planet model, and upon careful inspection of the spectral time-series data it became clear that OGLE-TR-33 was indeed a hierarchical triple system containing an eclipsing binary, and not a transiting planet at all—the apparent Doppler shift of the lines was actually an asymmetry in the combined line profiles of two stars. This and other similar examples were strong early warnings to the community that false positive scenarios may lurk behind any transit candidate, and that extremely careful analysis is required to rule out false positives in any individual case.

As a result, the traditional method of promoting a transit candidate signal to the status of a *bona fide* planet involves an in-depth, multi-instrument follow-up campaign. A prototypical example of this procedure is detailed in the discovery paper of TrES-1b (Alonso et al. 2004), which was the first transiting exoplanet to be discovered by a survey specifically designed for the purpose. The follow-up observations required to confirm this signal as a planet were the following:

- *H*- and *K*-band adaptive optics (AO) imaging to rule out contaminating stars of similar brightness

outside of  $\sim 0''.3$  separation from the target star.

- Medium-resolution Echelle spectroscopy (velocity resolution of  $8.5 \text{ km s}^{-1}$ ) at seven epochs, to rule out scenarios involving massive binary companions and to characterize the host star as suitable for further follow-up.
- Multi-color transit photometry with three different facilities in seven different filters (in addition to the discovery data), to check that the eclipse depth is color-independent, consistent with the eclipse of a cold, substellar object.
- High-precision ( $\sim 10 \text{ m s}^{-1}$ ) RV measurements with the Keck High Resolution Echelle Spectrometer (HIRES) at eight epochs, which detected the sinusoidal Doppler variation caused by the transiting planet.

The “funnel” procedure established with this discovery, in which a large number of candidates<sup>1</sup> are whittled down by various follow-up observations to identify the few that are worthy of observations with a large telescope such as Keck, was absolutely necessary in the early days of transit surveys, given the scarcity of precise RV spectrometers and the desire to discover and characterize individual exoplanetary systems. Throughout the last decade this has been widely adopted as the follow-up paradigm: find many candidates, identify a subset appropriate for follow-up, and eventually confirm a small number of these with RV measurements.

By the beginning of 2009, just before the launch of NASA’s new *Kepler* mission, transit surveys, having operated under this paradigm for about five years, had discovered 33 planets: a modest contribution compared to the  $>200$  planets that had by then been discovered by the radial velocity (RV) method. Still, these transiting planets—all at least half as massive as Jupiter—were of high scientific value because they had both mass and radius measured, fueling an industry of inquiry into understanding the radii of giant planets (Bodenheimer et al. 2003; Burrows et al. 2007; Seager et al. 2007; Fortney et al. 2007; Batygin & Stevenson 2010, e.g.). Demographic studies of exoplanets, however, were obliged to primarily use results from RV surveys, both because of the small number of available transiting planets and because of the messiness and various biases of the survey and confirmation processes.

## 2.2 The *Kepler* Mission

The *Kepler* spacecraft launched on March 7, 2009. The scientific goal of the mission was to conduct a survey for transiting exoplanets in order to determine the frequency of Earth-like planets around sun-like stars. As the signal of an Earth-like planet’s transit is a periodic dimming of about 100 parts per million, detecting such a miniscule signal requires photometric precision on the order of 10 parts per million—about  $100\times$  more precise than the very best ground-based photometry.

---

<sup>1</sup>Tres-1b was one of 16 transit-like signals among 12,000 stars monitored by the survey in its field and the only one to become positively identified as a planet.



To accomplish this ambitious task, the *Kepler* space telescope consists solely of a [1.2]m mirror and a focal plane paved with 42 CCDs. With no shutter, no filters, and no moving parts, it was built for a single purpose: to be the most precise photometer ever created. Its operational mode was also exceptionally simple: staring continuously at the same  $\sim 100$  square-degree field, saving pixel postage stamps of 150,000 chosen target stars every 30 minutes and an additional 500 stars with 1-minute cadence. Every three months the spacecraft would beam down that quarter’s worth of data and roll to keep its solar panels aligned properly, resulting in a 90-degree rotation of the field of view with respect to the camera. When the second of its four reaction wheels failed in May of 2013, the primary data-taking portion of the mission was over, since pointing could no longer be maintained to the level required for extreme photometric precision. During its four years of data collection, *Kepler* made [xxx] brightness measurements of [NNN] stars.

Before discussing the *Kepler* results in more detail, I will first give a general overview of the process by which a transiting planet candidate is identified by the *Kepler* team. The details and specific implementation of this process has evolved over the course of the mission, but the generic outline has remained the same. After reducing the pixel-level data to calibrated light curves for each target star, those light curves are searched for periodic dimmings; if the detection pipeline finds something it considers significant, it marks it a Threshold Crossing Event, or TCE. These are then investigated briefly by eye in a so-called “trriage” process intended to weed out obviously instrumental or clearly spurious signals from physical eclipses. TCEs that pass triage are then assigned a “*Kepler* Object of Interest” number such as KOI 1234.01. If there are multiple KOI signals in the light curve of the same target star, they are assigned sequential numbers like KOI 1234.02 and 1234.03.<sup>2</sup>

Next comes the more detailed review stage. Many KOIs can be identified as clearly being of stellar eclipse origin rather than due to a transiting planet, upon more detailed inspection of the *Kepler* photometry. For example, there may be a visible secondary eclipse inconsistent with a planetary interpretation. And in many cases, detailed analysis of how the center of light shifts during the transit signal can indicate that the source of the periodic dimming is actually a spatially offset blended eclipsing binary, that classic transit false positive. Indeed, the techniques that have been developed for this analysis are now sometimes capable of detecting eclipse sources that are as close as a few tenths of an arcsecond to the target star—especially impressive considering that the *Kepler* pixels measure 4” on a side. Various special circumstances such as saturation or nearby variable stars can cause systematic errors in these techniques, but they remain a powerful tool to vet KOIs for false positives. Once a KOI passes this inspection stage without any clear indication of being a false positive, it is promoted to “planet candidate” status.

The first public catalog of transiting planet candidates found in *Kepler* data was released in February 2011 (Borucki et al. 2011) and consisted of 1235 KOIs that were found in a search of the first  $\sim 120$  days of *Kepler* photometry (Q0-Q2). Since then there have been several other *Kepler* pipeline searches released: the first based on a search of Q1-Q6 data Batalha et al. (2012), and subsequent unpublished and still incomplete but

---

<sup>2</sup>Designations of signals in multi-KOI systems are made in the order they are discovered, which is usually in order of decreasing signal-to-noise ratio, and often *not* in order of increasing period.

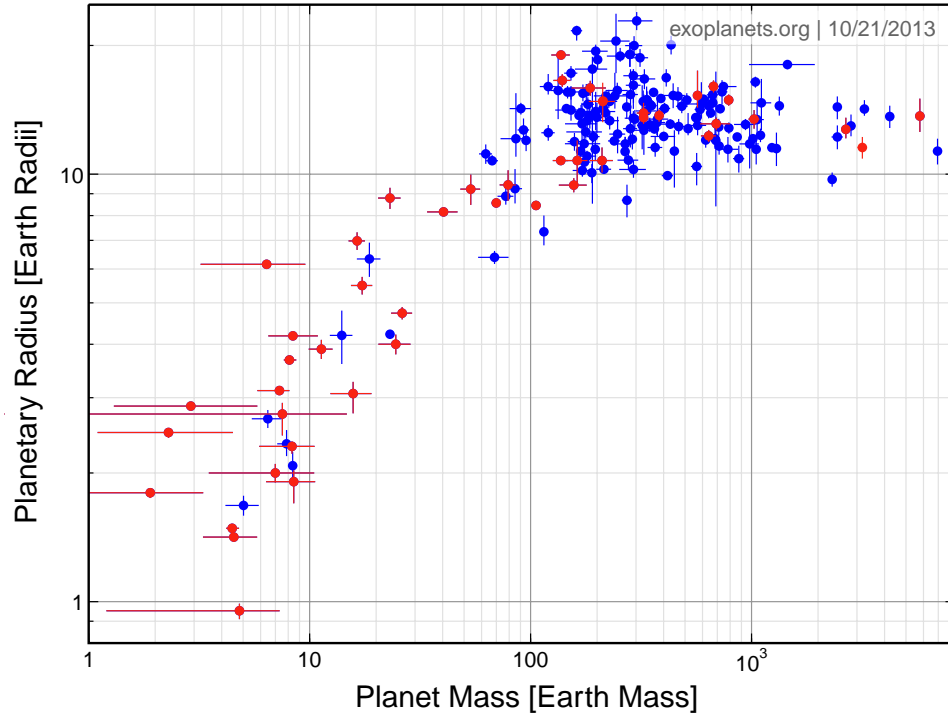


Figure 2.1 All exoplanets currently known with both masses and radii measured. *Kepler* discoveries (red points) dominate the low-mass, small-radius end of this distribution.

publicly released catalogs based on Q1-Q8 and Q1-Q12 data. These data products, as well as a cumulative table intended to host the latest information on each KOI, are currently being stored at the NASA Exoplanet Archive (NEA). The cumulative table contains 5779 KOIs, including 140 confirmed planets, 3448 planet candidates, and 2191 known false positives. The Q1-Q12 table, which contains the results of the most recent uniform search that is still in the vetting process, currently lists 850 candidates, 1338 false positives, and 2949 “not dispositioned” among a total of 5137 KOIs. The NEA also hosts a database of 18,406 TCEs.

Among the discoveries that *Kepler* has made are many exoplanet “firsts” and otherwise dramatic findings. The Kepler-9 planetary system was the first system discovered that had multiple transiting planets as well as the first to have masses measured via transit timing variations, or TTVs (Holman et al. 2010). Kepler-10b was the first rocky exoplanet to have a well-measured mass (Batalha et al. 2011).<sup>3</sup> The Kepler-11 planetary system has *six* transiting planets, five of which within the orbit of Mercury (Lissauer et al. 2011). Kepler-16b (dubbed “Tatooine” in the press) was the first confirmed circumbinary planet (Doyle et al. 2011). Kepler-22b was the first discovered transiting “habitable-zone” planet, meaning its estimated equilibrium temperature could plausibly allow for the presence of liquid water (Borucki et al. 2012). Kepler-20e and Kepler-20f were the smallest detected exoplanets at the time of their announcement (Fressin et al. 2012), but were soon supplanted by the planets of the Kepler-42 system, which are all three smaller than Earth, and all orbiting

<sup>3</sup>Corot-7b was the first rocky exoplanet detected, but its host star’s high level of activity made its confirmation challenging, with its mass measurement remaining uncertain.

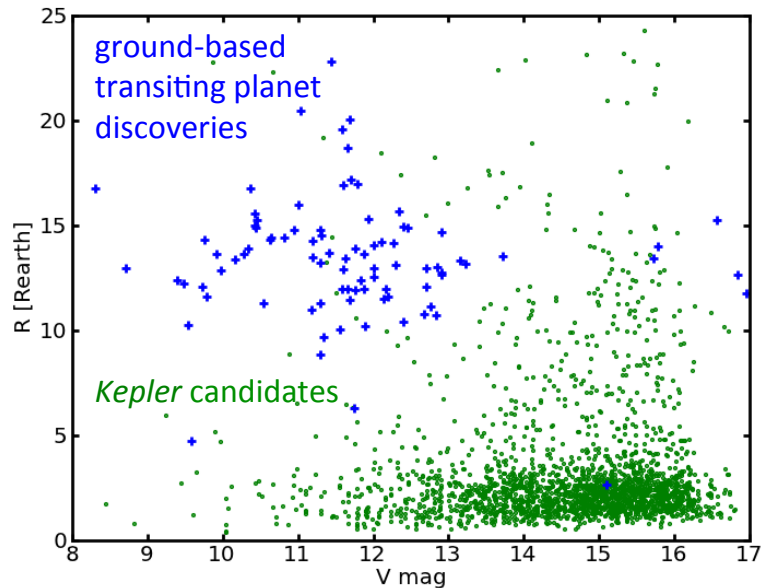


Figure 2.2 The difficulty following up *Kepler* candidates with the same strategies that have been used to follow up ground-based transit candidates. The vast majority of *Kepler* candidates are either too small to detect via RV measurements or are around host stars that make precise measurements too time-consuming.

within 0.02 AU of a  $0.13 M_{\odot}$  star Muirhead et al. (2012b)—similar to a slightly scaled-up system of Galilean satellites. Even this title was then surpassed by Kepler-37b, which is just barely larger than Earth’s moon (Barclay et al. 2013). There has also been a handful of super-Earth-sized *Kepler* planets whose masses have been measured via either RV measurements or TTVs—these planets have for the first time significantly populated the planet mass-radius diagram below the giant planet region (see Figure 2.2).

Despite all these headline-grabbing individual planetary systems, the core mission and true power of *Kepler* is statistical. While its stated primary goal is to determine the frequency of Earth-like planets, any effort that is able to accomplish this ambitious task will also reveal the demographics of the population of larger and shorter-period planets as well. However, standing between the *Kepler* discoveries and this goal is the as-yet unconfirmed nature of the thousands of planet candidates, and that traditional confirmation (as described for ground-based surveys in Section 2.1) is literally impossible for the vast majority of candidates given current and immediately foreseeable follow-up resources. Not only is the sheer number of candidates overwhelming (many more than have been followed up in the first decade of ground-based surveys), but the host stars are generally several magnitudes fainter than ground-based transit hosts, and the planets are mostly much smaller (Figure 2.2). While a number of *Kepler* candidates have indeed been confirmed with either RV observations or TTV measurements, the number of additional systems in which such confirmation is possible represents only a small minority of all *Kepler* candidates.

This central *Kepler* challenge has given birth to a new field in exoplanet research: probabilistic validation of transit candidates. The idea behind probabilistic validation is to demonstrate that the probability of a candidate signal to be a planet is overwhelmingly higher than the probability for it to be an astrophysical false positive. This concept was first pioneered by a technique known as BLENDER (Torres et al. 2011; Fressin et al. 2011, e.g.), but as this method is very time-consuming and resource-intensive, it has only been applied to a small number of candidates, and does not appear to be capable of efficiently analyzing the entire sample of KOIs. The research that I present in Chapters 5 and 6 of this thesis are the first contributions to this field equipped to handle the scale of the *Kepler* survey.

## Chapter 3

# Overview of Bayesian Inference

The goal of all observational science is to convert data into knowledge about the Universe. Sometimes this conversion is very straightforward; for instance, one may wish to know how many legs a centipede has, so one counts the centipede's legs and then knows that it has 42. Astronomy, however, is often more like trying to determine how many legs a centipede has, in a dark room, in under five minutes, using only a match and a low-resolution camera, when one is not even sure whether centipedes exist. And the room might or might not be crawling with millipedes. Divining knowledge from such scant, incomplete, or messy data requires the tools of statistics; using the mathematical language of probability to accomplish this is known as Bayesian inference. As much of the work in this thesis employs Bayesian inference in some form, I here give it a concise introduction.

All of Bayesian inference is built on what is known as Bayes' equation, which is in turn derived directly from the concept of conditional probability. Conditional probability is best described with an example. The probability to obtain snake eyes (two "1" rolls) in a toss of two six-sided die is 1 in 36. However, if the dice are rolled sequentially and the first displays a "1," then the probability to obtain snake eyes becomes 1 in 6: the probability for both dice to show "1" *given that one is already known to do so* is the same as the probability of the single remaining die to show "1." Mathematically, conditional probability may be defined relative to the joint probability of two events:

$$P(A \text{ and } B) = P(A|B)P(B), \quad (3.1)$$

where the vertical bar is read "given." Since  $P(A \text{ and } B) = P(B \text{ and } A)$ , that means

$$P(A|B)P(B) = P(B|A)P(A), \quad (3.2)$$

which leads directly to Bayes' equation:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad (3.3)$$

I consider Bayesian analysis to be any statistical inference that uses this equation as a starting point.

### 3.1 Parameter Estimation

The most common application of Bayesian inference is to quantify precisely how a data set provides knowledge about the true values of the parameters of a specific model. A “model” in this case means an underlying theory (either physically motivated or just descriptive) that predicts the value of data as a function of one or more parameters. A simple example of this is time series data that is expected to follow a linear trend. The quantity being measured—e.g., flux  $F$  from an astronomical source at time  $T$ —can thus be described as

$$F = aT + b. \quad (3.4)$$

One can imagine therefore making flux measurements of this source with the goal of trying to determine the value of the parameters  $a$  and  $b$ .

This may be accomplished by applying Bayes’ equation. Consider rewriting Equation 3.1 using *probability density functions* (PDFs) of random variables  $X$  and  $Y$  rather than probabilities of discrete events  $A$  and  $B$ :

$$f_{XY}(x, y) = f_{XY}(x|Y = y)f_Y(y). \quad (3.5)$$

Note that the conditional probability distribution in this continuous case is defined as the joint probability distribution of both variables with the value of one of the variables fixed. Equation 3.3 then becomes

$$f_{XY}(x|Y = y) = \frac{f_{XY}(y|X = x)f_X(x)}{f_Y(y)}. \quad (3.6)$$

$X$  and  $Y$  may now be replaced by “model” and “data,” and some notation changed for the more familiar form of Bayes’ equation to emerge:

$$p(\text{model}|\{\text{Data}\}) = \frac{\mathcal{L}(\{\text{Data}\}|\text{model})\pi(\text{model})}{\text{Ev}(\text{data})}. \quad (3.7)$$

Here is where some typical terminology may be introduced. The probability distribution  $p$  on the LHS of Equation 3.7 is known as the **posterior probability distribution** for the model parameters,  $\mathcal{L}$  is known as the **likelihood**, and  $\pi$  is known as the **prior**. The denominator on the RHS is the **Bayesian evidence** (sometimes called **marginalized likelihood**). I will discuss this factor in detail in Section 3.2; for current purposes, it is simply a normalization constant.

Returning to our example then, the goal is to calculate the posterior probability distribution of the model parameters  $a$  and  $b$  given the observed data, which is a series of flux measurements  $F_i$  taken at times  $T_i$  with measurement uncertainties  $\sigma_i$ . This means evaluating the RHS of Equation 3.7, which in turn requires

expressions for the likelihood and prior functions. As long as the data points are independent of each other, the likelihood function may first be factored into an individual likelihood function for each data point:

$$\mathcal{L}(a, b | \{F, \sigma\}) = \prod_i \ell_i(a, b | F_i, \sigma_i). \quad (3.8)$$

The form of each of these individual likelihoods  $\ell_i$  depends on the same assumptions about the data that go into determining the error bars—from what probability distribution do we expect the data to be drawn? For astronomical data (such as photometry) this is often a Poisson distribution, usually with a large enough mean that it is very well described by a normal distribution. For our example then, the likelihood of each data point is

$$\ell_i = \exp \left( \frac{-(F_i - (aT_i + b))^2}{2\sigma_i^2} \right), \quad (3.9)$$

a normal distribution with mean and width defined by the model prediction and the measurement uncertainty, evaluated at the observed values, as illustrated in Figure 3.1.

The prior probability distribution on the parameters is intended to represent whatever is known about the model parameters before any data was taken. Often it is preferable to use “ignorant” priors, which typically means a uniform or log-uniform probability distribution within some plausible range. Occasionally, however, there may exist an independent measurement of one of the parameters in question, and so the prior may encapsulate the results of that measurement rather than claim ignorance. For our example, we will use a uniform prior for both of our parameters, between -5 and 5 for the slope parameter  $a$  and between 0 and 20 for the y-intercept parameter  $b$ , leading to the following trivial joint prior PDF:

$$\pi_{ab}(a, b) = \frac{1}{10} \cdot \frac{1}{20} = \frac{1}{200} \quad (3.10)$$

With the likelihood and prior defined, we now have enough information to calculate the joint posterior probability distribution for parameters  $a$  and  $b$ , conditioned on our observed data. Taking a grid of parameter values in our prior range, we evaluate the product of the likelihood and prior at each point in the grid, with the results depicted in Figure 3.1. The contours represent the regions of parameter space encompassing 68%, 95%, and 99% of the total posterior mass, and the ‘+’ indicates where the posterior attains a maximum, indicating the “best fit” value of the parameters. On the top and side of this figure are illustrated the “marginalized” posterior of each parameter, obtained by integrating out the other:

$$p_a(a) = \int_0^2 p_{ab}(a, b) db \quad ; \quad p_b(b) = \int_{-5}^5 p_{ab}(a, b) da \quad (3.11)$$

It is worth noting that what this section has described so far is at face value no different from parameter estimation by traditional least-squares optimization, or “minimizing  $\chi^2$ ,” as the logarithm of the likelihood in

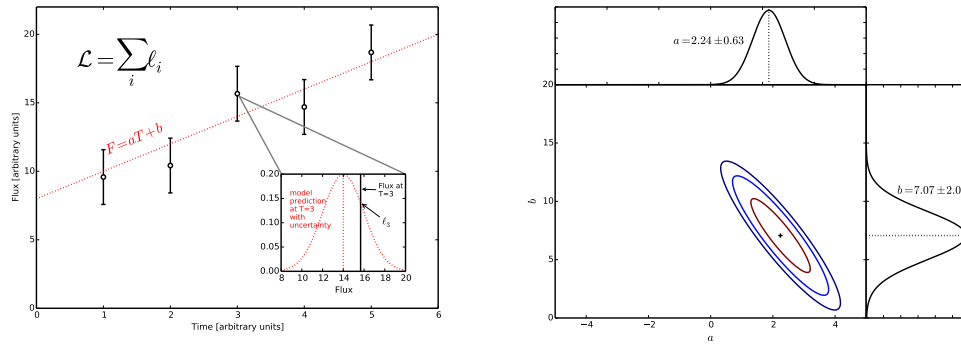


Figure 3.1 Demonstration of Bayesian parameter estimation. The left panel shows a linear model being tested against observed data. The total likelihood of the data  $\mathcal{L}$  is the product of all the individual likelihoods  $\ell_i$ , each of which is determined by evaluating a likelihood function (inset, red-dotted) at the value of the observed data point. In this example each individual likelihood function is a normalized Gaussian centered at the model value predicted for that time, whose width is determined by the observational uncertainty for that point. The right panel shows the posterior probability distribution of the model parameters  $a$  and  $b$ , along with the marginalized distribution for each. The contours enclose 68%, 95%, and 99% of the total probability.

Equation 3.8 is exactly  $-\frac{1}{2}\chi^2$ , where

$$\chi^2 = \sum \frac{(F_i - (aT_i + b))^2}{\sigma_i^2} \quad (3.12)$$

Minimizing the  $\chi^2$  function with respect to the parameters  $a$  and  $b$  gives the exact same best-fit parameter solution as does maximizing the posterior, because the two are mathematically identical. There are, however, many advantages that this probabilistic formulation of parameter estimation has over simple least-squares optimization:

- **Priors.** Maximizing the posterior is only equivalent to minimizing  $\chi^2$  when priors on the parameters are strictly uniform. This means that a uniform prior is always an implicit assumption in least-squares minimization. Often Bayesian analysis is criticized for its use of priors; it is thus important to realize that there are always priors, and Bayesian inference just makes them explicit (and allows them to be flexible).
- **Non-Gaussian Likelihoods.** Once again, the  $\chi^2$  formulation carries the implicit assumption that the likelihoods of the observed data are strictly Gaussian, which is not always the case. Occasionally likelihoods will follow a different distribution, such as Poisson or binomial counting statistics, depending on the nature of the observation. A likelihood may even be based on a completely non-standard probability distribution constructed via a complex Monte Carlo simulation—Chapters 4, 5, and 6 all use this technique. There is nothing analogous to this in standard least-squares analysis.
- **Posterior Uncertainties.** Perhaps the greatest advantage of Bayesian parameter estimation is its ability to precisely quantify the full posterior probability distribution of model parameters. Formal uncertainties that are provided by least-squares minimization or similar techniques rest on the assumption that



not only are the likelihood functions Gaussian, but also that the posterior is a normal distribution (a necessary consequence of assuming Gaussian likelihood functions and flat priors in a linear model). Often, however, this is not correct, such as when there are complicated parameter correlations in a non-linear model or when non-uniform priors play an important role.

One final note on parameter estimation: when a model has more than two or three parameters, evaluating the likelihood function on a parameter grid quickly becomes computationally infeasible. However, there are several sampling techniques—Markov Chain Monte Carlo (MCMC) being the most widely used—that can reliably reconstruct the important features of the posterior without having to exhaustively explore the entire prior range. This enables efficient and straightforward parameter estimation with reliable uncertainty or correlation determination even for high-dimensional models.

## 3.2 Model Selection

One limitation of parameter estimation is that while one may precisely calculate a posterior distribution for the parameters of a model, there is no way to judge whether the model itself is even a sensible model to be using; that is, there is no “goodness-of-fit” metric. For example, one may use MCMC to fit a Keplerian orbit to an RV time series and derive posteriors on the model parameters, but that in itself would in no way determine whether it is justified to claim that a signal is even present. However, there is a Bayesian way to address this.

Model selection is intended to answer the question “is model A or model B a more appropriate model to use to describe this data set?” One of the subtleties in answering a question like this is the problem of overfitting. Looking back at the mock data set from §3.1, one might imagine that a higher-order polynomial might fit the data better; that is, the deviations of the data from the model could be smaller in the least-squares sense with a quadratic or cubic. In fact, with a fourth-order polynomial, the fit would be exact, though completely useless—there would be no information extracted from the data at all. The central question to ask in this case is whether a more complex model is *justified* or not: does the better fit of a higher-order polynomial make up for its added complexity? Or, returning to the example of the RV time series, is a Keplerian model justified over a null model such as a simple linear fit?

The Bayesian way of answering this question goes back, as always, to Bayes’ equation. Now, rather than considering the probability distribution of parameters in a model, consider the probability of the following hypothesis: “model A is the correct model to describe reality.” Call this hypothesis  $H_A$ , and Bayes’ equation gives the following expression for the posterior probability of  $H_A$ , given the observation of data  $D$ :

$$\Pr(H_A|D) = \frac{\Pr(D|H_A)\Pr(H_A)}{\Pr(D)}. \quad (3.13)$$

The posterior probability of  $H_B$ , the hypothesis that model B is correct, is exactly analogous, and so the ratio

of probabilities of the two hypotheses, or *odds ratio*, is the following:

$$\frac{\Pr(H_A|D)}{\Pr(H_B|D)} = \frac{\Pr(D|H_A)\Pr(H_A)}{\Pr(D|H_B)\Pr(H_B)}. \quad (3.14)$$

The  $\Pr(D)$  factor cancels in the ratio. What are these factors on the RHS of this equation that must be evaluated to determine this odds ratio?  $\Pr(D|H_A)$  and  $\Pr(D|H_B)$  are serving the function of the likelihood from the nomenclature of Equation 3.7, whereas  $\Pr(H_A)$  and  $\Pr(H_B)$  function as priors. To try to avoid confusion, I will call these for now **hypothesis likelihoods** and **hypothesis priors**.

It seems at first glance that little progress has been made; we have re-written Bayes' equation and now we just have new likelihoods and priors to worry about. However, consider now what the version of Bayes' equation that would be used for parameter estimation under hypothesis  $A$  looks like, where  $\theta_A$  is the parameter vector used in model  $A$ :

$$\Pr(\theta_A|D) = \frac{\Pr(D|\theta_A)\Pr(\theta_A)}{\Pr(D)}. \quad (3.15)$$

Now note that it is implied in Equation 3.15 that it is also “given” that we are assuming model  $A$  is the correct model (even though we currently make no judgment whether that is appropriate or not); otherwise, solving for the parameters of model  $A$  would be nonsensical. We can therefore make this additional condition explicit, without changing anything about what is being calculated, and write the following:

$$\Pr(\theta_A|D, H_A) = \frac{\Pr(D|\theta_A, H_A)\Pr(\theta_A|H_A)}{\Pr(D|H_A)}. \quad (3.16)$$

The denominator here is exactly what we described above as the “hypothesis likelihood” for model  $A$  from Equation 3.14, and because we know that its function in this context must be as a normalization factor (ensuring that both sides integrate to unity) we know how to calculate it:

$$\Pr(D|H_A) = \int_{\Theta_A} \Pr(D|\theta_A, H_A)\Pr(\theta_A|H_A)d\theta_A, \quad (3.17)$$

where the integral is over the whole parameter space  $\Theta_A$ . In other words, this “hypothesis likelihood” factor is equivalent to the integral of the product of the “parameter likelihood” and “parameter prior” over the whole parameter space of the model hypothesized. Returning to the notation of Equation 3.7, we can thus say that

$$\frac{\Pr(H_A|D)}{\Pr(H_B|D)} = \frac{\pi_{H_A}}{\pi_{H_B}} \cdot \frac{\int_{\Theta_A} \mathcal{L}_A(\theta_A|D)\pi_A(\theta_A)d\theta_A}{\int_{\Theta_B} \mathcal{L}_B(\theta_B|D)\pi_B(\theta_B)d\theta_B} \equiv \frac{\pi_{H_A}}{\pi_{H_B}} \cdot \frac{\text{Ev}(A)}{\text{Ev}(B)}. \quad (3.18)$$

This is Bayesian model selection. If there are two models competing to explain a given data set, the odds ratio of the two hypothesis is the ratio of their “evidences,” as here defined, multiplied by the ratio of the hypothesis priors, meaning the relative probability of the two models that would have been predicted before

the gathering of any data. In other words, the probability that model  $A$  is the correct model is

$$\Pr(H_A) = \frac{\pi_{H_A} \text{Ev}(A)}{\pi_{H_A} \text{Ev}(A) + \pi_{H_B} \text{Ev}(B)}. \quad (3.19)$$

This idea may be extended to considering more than one hypothesis, in which case

$$\Pr(H_i) = \frac{\pi_{H_i} \text{Ev}(i)}{\sum_i \pi_{H_i} \text{Ev}(i)}. \quad (3.20)$$

The most important thing to understand about Bayesian model selection is that one can compare models regardless of their relative complexity, without worry that a more complicated model will be favored due to overfitting. The reason for this is that calculation of the evidence has the so-called “Occam’s razor” built in that will naturally prefer simpler models: if there is a very large prior range or many parameters, the height of the prior function will be significantly lower than if there is a smaller parameter space. Thus, in order for a more complicated model to win, the likelihood must be correspondingly higher than in the simpler model—in other words, it must be a much better fit.

Actually calculating these evidence factors is often much more challenging than simply writing down their definitions; in fact, with models with more than just a few parameters, it becomes nearly impossible to calculate the entire integral, for the same reason that it becomes intractable to calculate the likelihood on a fine grid in a high-dimensional parameter space. However, there have been advanced computational methods developed, such as nested sampling (Feroz et al. 2009), that can efficiently accomplish this difficult task, even for large numbers of parameters.

### 3.3 Final Thoughts on Bayesian Inference

In this section I have separated discussion of parameter estimation and model selection, which is perfectly appropriate because they are naturally separable concepts and are nearly always treated separately. However, I argue that there is fundamentally no difference between the two endeavors, rooted as they both are directly in Bayes’ equation. In fact, I would advocate for the obfuscatory term “evidence” to be replaced by the much more descriptive “hypothesis likelihood,” a term which I have actually never seen used before in this context. However, I believe that this new term would make clear the explicit connection that model selection has with parameter estimation.

Consider that Equation 3.20 is structurally identical to Equation 3.7, the only difference being that Equation 3.20 uses discrete index notation whereas Equation 3.7 consists of continuous functions. Other than that, they are identical. Conceptually, one might say that parameter estimation is just the continuous limit of discrete model selection between many different models, where the difference between models is just the value of the parameters. On the other hand, one might equally say that model selection is just parameter

estimation where the parameter being estimated is a “model index”  $i$ , and where the likelihood and prior functions change discontinuously as a function of  $i$ .

I share these thoughts in the introductory section of this dissertation because in the published papers where I have applied these concepts (Chapters 4, 5, and 6), I have not always used standard terminology; for example, in Chapters 5 and 6 I use  $\mathcal{L}$  instead of “evidence,” even though the factor in question is exactly what I call here the hypothesis likelihood. I also do not explicitly emphasize in those papers that I am doing justifiably rigorous Bayesian model selection even though upon reflection I believe that I am.

In any case, Bayesian analysis is relatively new to astronomy, and this introduction is intended to give a clear and concise overview of the subject, as well as to provide technical background for the research I present in the following chapters.

## **Part II**

# **Original Research**

## Chapter 4

# Discerning Exoplanet Migration Models Using Spin-Orbit Measurements

*Before Kepler and its revolutionary results began dominating the exoplanet headlines, there was considerable attention given to the surprising fact that many short-period giant planets were being discovered to have retrograde or highly inclined orbits relative to the spins of their host stars. This work was my first foray into exoplanet demographics and was published as “Discerning Exoplanet Migration Models Using Spin-Orbit Measurements,” ApJ 729,138 (2011), by TDM and John Johnson. While the study of exoplanet migration and spin-orbit alignment has progressed significantly since the publication of this work, it remains an illustrative example of how to apply Bayesian model selection to use observed data to distinguish between different theories.*

### 4.1 Introduction

Exoplanets that transit their host stars provide opportunity to study distant planetary systems in great detail. Most immediately, photometry during transit measures a planet’s radius and density, but follow-up studies can provide much more. Secondary eclipse photometry, photometric phase curve measurements, and transmission spectroscopy, for example, can reveal a planet’s temperature, albedo, atmospheric composition, and even weather patterns. While these tools investigate the physical characteristics of the planets themselves, transits also provide a valuable opportunity to measure details of planets’ orbital characteristics using the Rossiter-McLaughlin (RM) effect.

The RM effect is an anomalous Doppler signal due to the shadow of a transiting planet crossing the face of a rotating star, and is measured by obtaining radial velocity (RV) measurements of the host star during transit. As the approaching limb of the stellar surface is occulted, the total integrated radial velocity is redshifted, and as the receding limb is occulted, the integrated velocity is blue-shifted. Modeling the RM effect yields a measure of the angle between the orbital axis of the planet and the projected rotation axis of the host star, typically referred to as  $\lambda$ . While this angle is not itself physically meaningful, it constrains  $\psi$ , the true angle

between the planet’s orbit and the star’s rotation. In addition to being a fundamental system parameter akin to semimajor axis or eccentricity,  $\psi$  is a potential window into learning about planetary orbital migration, as different migration scenarios predict different distributions of  $\psi$ .

The first several RM measurements that were made all indicated small values of  $\lambda$  (Winn et al. 2005, 2006; Wolf et al. 2007; Narita et al. 2007), and thus were consistent with small values of  $\psi$ , which was not unexpected since the orbits of all the planets in the Solar System are aligned to within  $7^\circ$  of the solar spin axis. The first misaligned system, XO-3, was discovered by Hébrard et al. (2008), and confirmed by Winn et al. (2009a). Since then, many additional misaligned systems have been discovered (Winn et al. 2009b; Johnson et al. 2009; Narita et al. 2009; Triaud et al. 2010). This diversity of measured  $\lambda$ s suggests that planetary migration is more complicated than simple disk migration, which predicts planet orbits well aligned with stellar spins (unless the disk itself is misaligned), and even hints at multiple migration channels.

As the number of exoplanet systems with measured values of  $\lambda$  increases, so does the desire and ability to draw conclusions based on the data. For example, Fabrycky & Winn (2009) (FW09) suggest that there might be two distinct populations of close-in planets—intrinsically aligned and intrinsically misaligned—with a 95% probability of  $> 64\%$  of planets belonging to the aligned population. This remarkable result was based on the first 10 RM measurements, which included a single misaligned system.

There now exists a sample of 32 published spin-orbit angles<sup>1</sup>, which provides a valuable opportunity to readdress and extend this previous study, particularly in light of the recent proposition by Triaud et al. (2010) that current data suggest that all hot Jupiters migrated via the Kozai mechanism (Kozai 1962; Wu et al. 2007; Fabrycky & Tremaine 2007). The central goal of this paper is to determine whether the current sample of spin-orbit measurements is sufficient to begin to discern between the predictions of different exoplanet migration theories such as the Kozai mechanism and planet-planet scattering (Nagasawa et al. 2008), and if not, then how many more RM observations will be needed to draw more meaningful conclusions about the intrinsic  $\psi$  distribution of transiting exoplanets.

We describe our models in §4.2 and our analysis in §4.3 and §4.4. In §4.5 we repeat our analysis restricted to hot stars (following the suggestion of Winn et al. (2010)). In §4.6, we look to the future and ask how many RM observations will be needed to draw more confident statistical conclusions. We conclude our discussion in §5.5.

## 4.2 The Models

We test two hypotheses against each other in this paper. The first is that close-in exoplanets migrated to their present-day orbital locations through a combination of the effects of Kozai cycles and tidal friction, as described by Fabrycky & Tremaine (2007) (FT07). Kozai cycles are oscillations in eccentricity and inclination of a close binary system caused by the presence of distant third companion in an inclined orbit. If a giant

---

<sup>1</sup>As of publication of this article, in ApJ 729,138 (2011).

planet (assumed to have formed beyond the ice line in its protoplanetary disk) undergoes Kozai cycles that cause it to pass within a few stellar radii of its host star, then tidal friction can quickly circularize its orbit, freezing in a potentially large orbital inclination ( $\psi$ ) to the newly-migrated hot Jupiter. FT07 uses 1000 simulations of such systems, using Jupiter-mass planets with initial orbital separations of 5 AU and outer  $1 M_{\odot}$  companions on 500 AU orbits highly inclined with respect to the planet’s initial orbit. The final inclination of the resulting close-in planets in these simulations is their prediction of the hot Jupiter  $\psi$  distribution resulting from this migration mechanism (FT07, Figure 10). In the rest of the paper, we refer to this model as “KCTF.”

The second hypothesis is that planet-planet scattering is the dominant mechanism for forming close-in planets, as modeled by Nagasawa et al. (2008) (N08). In their simulations, they study the evolution of systems of three Jupiter-mass planets with initial orbital separations of 5.00, 7.25, and 9.50 AU and initial inclinations of  $0.5^{\circ}$ ,  $1.0^{\circ}$ , and  $1.5^{\circ}$ . In addition to planet-planet scattering, they also include the effects of the Kozai mechanism and tidal friction, and produce a final population of close-in planets with a large range of orbital inclinations (N08, Figure 11(c)). We refer to this model as “PSTF.”

There have been many different simulations similar in nature to these that we test (e.g. Wu et al. (2007); Chatterjee et al. (2008); Ford & Rasio (2008); Jurić & Tremaine (2008)), but these are the two that produce the broadest distribution of  $\psi$  values, and thus seem most likely to be able to explain the observed retrograde systems. In addition, Triaud et al. (2010) also compared the observed data with these two models, concluding that the KCTF model of FT07 describes the current data better than PSTF of N08.

We emphasize that while the true spin-orbit angle  $\psi$  is the physically meaningful angle, only the projected version of this angle  $\lambda$  is measurable through the RM effect, because the inclination of the stellar rotation axis is unknown. If the star is rotating completely edge-on ( $I_{\star} = 90^{\circ}$ ), then  $\lambda = \psi$ , but in general the star’s rotation axis may be tilted along the line of sight, resulting in  $\lambda \neq \psi$  (see Fig. 3 of FW09). In other words, observation of a large value of  $\lambda$  is firmly indicative of a large  $\psi$ , but observation of a small  $\lambda$  does not rule out the possibility of large  $\psi$ . In rare cases, this unknown stellar inclination can be constrained by combining photometric determination of a rotation period with an estimate of the stellar radius and rotational line broadening (Winn et al. 2007). Additionally, the likelihood for any particular transiting planet to be misaligned may be estimated before even any RM measurement occurs, by comparing the observed line broadening to theoretical rotation predictions (Schlaufman 2010). In general however, since the inclinations of individual stellar rotations are unknown, statistical methods must be employed to draw conclusions about  $\psi$  from an observed population of  $\lambda$ .

One strategy to do this, employed by Triaud et al. (2010), is to statistically deproject each  $\lambda$  measurement to form a posterior probability density function (PDF) for  $\psi$  for each system, assuming an isotropic distribution for the inclination of stellar spins, and sum them to form an overall  $\psi$  PDF. However, this method has the disadvantage that isotropy of stellar spins is actually a strong assumption, since the observed distribution of spin inclinations will depend on the true  $\psi$  distribution, which we are trying to determine. This is analogous to how the orbit inclinations of RV-detected planets may not always be assumed to be isotropically



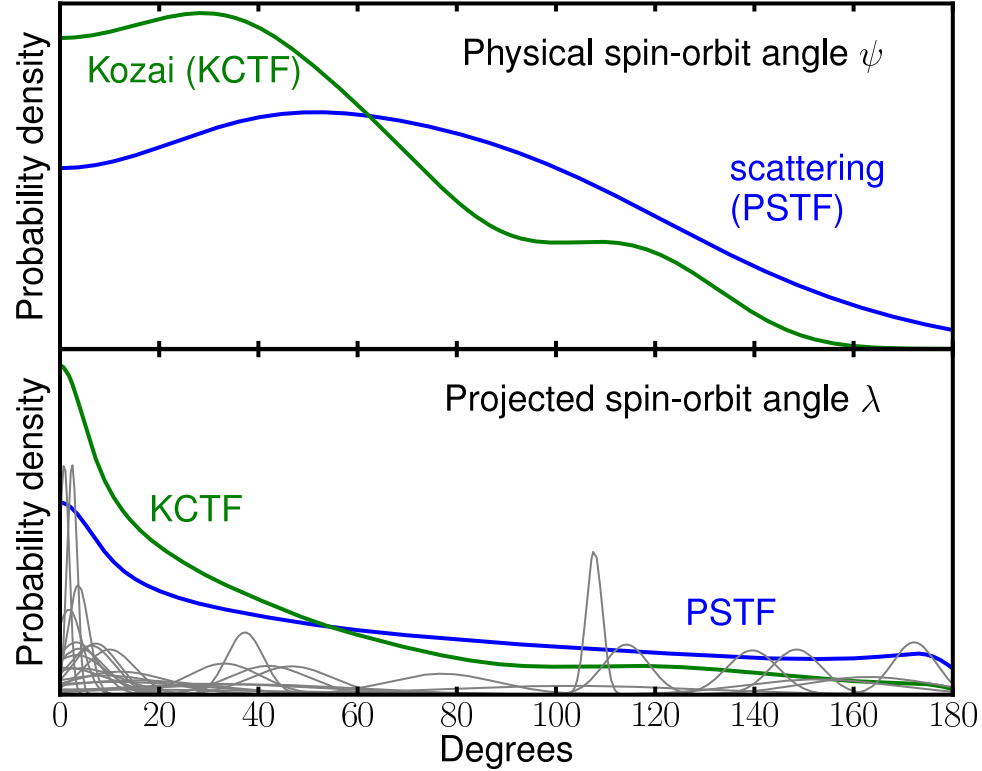


Figure 4.1 Probability distributions for the true ( $\psi$ ; top panel) and projected ( $\lambda$ ; bottom panel) spin-orbit angles that our two misalignment mechanisms produce. The KCTF  $\psi$  distribution is taken from the simulations of final inclinations of planets migrating through Kozai cycles + tidal damping from Fabrycky & Tremaine (2007), and the planet-planet scattering (PSTF) prediction is taken from the simulations of Nagasawa et al. (2008). We use Monte Carlo simulations to project each  $\psi$  distribution to create the  $\lambda$  distributions. Note that non-zero but small  $\lambda$  values are preferred by the KCTF model, while large values ( $\lambda > 60^\circ$ ) are preferred by the PSTF model. The bottom panel shows the distribution of the 32 measured  $\lambda$  values.

distributed, as discussed in Ho & Turner (2010a).

Our analysis relies instead on comparing the observed  $\lambda$  data directly to theoretical predictions of the models, in  $\lambda$ -space. This requires that we determine a probability density function (PDF) for  $\lambda$  corresponding to each of the  $\psi$  distributions we wish to test. Since both of our  $\psi$  predictions are the results of complicated simulations, we use a Monte Carlo procedure to perform this transformation.

We fix an observer-oriented spherical coordinate system and simulate a large number of systems as follows. First, we populate  $10^6$  stars with transiting planets with orbital inclination vectors  $\vec{O}$  assigned according to the distribution of known transiting planet inclinations. Then the stellar spins  $\vec{S}$  are assigned relative to the planet population, according to the predicted  $\psi$  distributions of KCTF and PSTF. This is accomplished by selecting a  $\psi$  from the distribution implied by the simulations of FT07 or N08, treating this  $\psi$  as a polar angle relative to  $\vec{O}$ , and assigning  $\vec{S}$  to have an azimuthal angle around  $\vec{O}$ , chosen uniformly on  $(0, 2\pi)$ . Then  $\vec{S}$  is

transformed back into the observer-oriented coordinate system, in which  $\lambda$  is simply the azimuthal angle of  $\vec{S}$  about the line of sight. The probability distribution for  $\lambda$  is then determined from the distribution of these resulting  $\lambda$  values. Figure 4.1 illustrates the original  $\psi$  and derived  $\lambda$  distributions for both misalignment models, as well as the current  $\lambda$  data.

In addition to comparing these two misalignment models, we also consider the possibility that there might be multiple migration mechanisms, inspired by the conclusions of FW09, who found the data available at the time (10 systems) favored a model with planets drawn from two distinct distributions: one perfectly aligned ( $\psi = 0$ ) and one isotropically misaligned. After introducing our methods in §4.3, we explore in §4.4 what we can learn if we assume only a fraction  $f$  of systems are misaligned according to one of the above mechanisms, with the remaining  $1 - f$  fraction being perfectly aligned.

### 4.3 Data Analysis

The goal of our analysis is to select the model that describes the data best, and to determine the confidence with which we can make that selection. We do this first assuming that all planets are misaligned, and then again allowing for an aligned population. The following sub-sections outline the details of these steps. The data we use are the 28 RM measurements compiled in Table 1 of Winn et al. (2010) (with 5 of these systems updated; Winn, in prep), plus HAT-P-14 ( $185^\circ \pm 4.5$ ), (Winn, in prep), HAT-P-4 ( $-15^\circ \pm 16$ ) (Winn, in prep), and XO-4 (Narita et al. 2010). A summary of the results of all our analysis is in Table 1.

#### 4.3.1 Which misalignment mechanism is preferred?

We use the Bayes factor for our model selection statistic, which in the simple case of comparing two models with no free parameters reduces to a likelihood ratio, as follows:

$$\mathcal{R} = \log_{10} \left( \frac{\mathcal{L}_{\text{Koz}}(\{\lambda\})}{\mathcal{L}_{\text{scat}}(\{\lambda\})} \right), \quad (4.1)$$

where  $\mathcal{L}_{\text{Koz}}(\{\lambda\})$  and  $\mathcal{L}_{\text{scat}}(\{\lambda\})$  are likelihoods of the observed data  $\{\lambda\}$  under the two different misalignment models.  $\mathcal{R} > 0$  favors the KCTF model, and  $\mathcal{R} < 0$  favors the PSTF model. The likelihoods are calculated as follows:

$$\mathcal{L}_{\mathcal{M}}(\{\lambda\}) = \prod_{i=1}^N P_{\mathcal{M}}(\lambda_i), \quad (4.2)$$

where

$$P_{\mathcal{M}}(\lambda_i) = \int_0^{180} p_i(\lambda) p_{\mathcal{M},\lambda}(\lambda) d\lambda, \quad (4.3)$$

with  $p_i(\lambda)$  being the probability distribution of the  $i$ th  $\lambda$  measurement (which we take to be a Gaussian centered at the published value with width as the published error bar), and  $p_{\mathcal{M},\lambda}(\lambda)$  being the  $\lambda$  PDF for the model in question ( $\mathcal{M} = \{\text{Koz}, \text{scat}\}$ ).

Using the current set of 32  $\lambda$  measurements, we calculate  $\mathcal{R} = 1.34$ , which favors KCTF over the PSTF. The following section explains how we quantify the strength of this model selection.

### 4.3.2 Confidence Assessment

Within the dichotomous paradigm of comparing two misalignment models we can determine the confidence in our model selection by answering the following question: “Given a measured value of  $\mathcal{R} = 1.34$ , which favors the KCTF model, what is the probability that the KCTF model is actually correct?” Or, more generally, given any measured value of  $\mathcal{R}$ , how can we quantify the confidence in the implied model selection?

To address this question, we perform the analysis described above on many simulated data sets. Starting with  $10^5$  randomly-generated spin-orbit systems for each misalignment model, constructed according to the procedure described in §4.2, we randomly select 32  $\lambda$  values from each underlying model. Each simulated  $\lambda$  measurement is the exact value of  $\lambda$  drawn from the simulations, perturbed by a measurement error  $\sigma_\lambda$ , assigned using Eq. 16 from Gaudi & Winn (2007):

$$\sigma_\lambda = \frac{\sigma}{v \sin I_* \sqrt{N}} \frac{1 - \gamma^2}{\gamma^2} \left[ \frac{(1 - b^2) \cos^2 \lambda + 3b^2 \cos^2 \lambda}{b^2(1 - b^2)} \right]^{1/2}, \quad (4.4)$$

where  $\sigma$  is the single-point RV measurement uncertainty,  $v \sin I_*$  is the projected rotational velocity of the star,  $N$  is the number of RV points in transit,  $\gamma$  is the planet-star radius ratio, and  $b$  is the transit impact parameter. For these simulated data sets, we randomly assign  $v \sin I_*$ ,  $\gamma$ , and  $b$  by drawing randomly with replacement from the current set of all transiting planet systems<sup>2</sup>. We take  $\sigma = 5$  m/s and  $N = 20$  for each simulated measurement.

We repeat this data simulation procedure 5000 times and calculate  $\mathcal{R}$  (Eqn. 4.1) for each data set, giving us an understanding of the expected distribution of  $\mathcal{R}$  if either of these models do describe the actual underlying  $\psi$ . Using these simulations to construct PDFs for  $\mathcal{R}$  under each model ( $p_{\mathcal{R}, \text{Koz}}(\mathcal{R})$  and  $p_{\mathcal{R}, \text{scat}}(\mathcal{R})$ ) we can then ask what the probability is of either model being true, given a measurement of  $\mathcal{R}$ . Applying Bayes’ theorem with a uniform prior on which model should be correct, we may write:

$$\Pr(\mathcal{M}|\mathcal{R}) = \frac{\Pr(\mathcal{R}|\mathcal{M})}{\sum_{\mathcal{M}} \Pr(\mathcal{R}|\mathcal{M})} = \frac{p_{\mathcal{R}, \mathcal{M}}(\mathcal{R})}{\sum_{\mathcal{M}} p_{\mathcal{R}, \mathcal{M}}(\mathcal{R})}, \quad (4.5)$$

where again  $\mathcal{M} = \{\text{Koz}, \text{scat}\}$ . For our specific case, this becomes:

$$\Pr(\text{Koz}|\mathcal{R} = 1.34) = \frac{p_{\mathcal{R}, \text{Koz}}(1.34)}{p_{\mathcal{R}, \text{Koz}}(1.34) + p_{\mathcal{R}, \text{scat}}(1.34)} = 0.94, \quad (4.6)$$

giving 94% confidence in the KCTF model. Figure 4.2 illustrates this confidence-assessment procedure.

This result appears to support the conclusion of Triaud et al. (2010), who claim that the current RM

---

<sup>2</sup>as listed at [www.exoplanets.org](http://www.exoplanets.org)

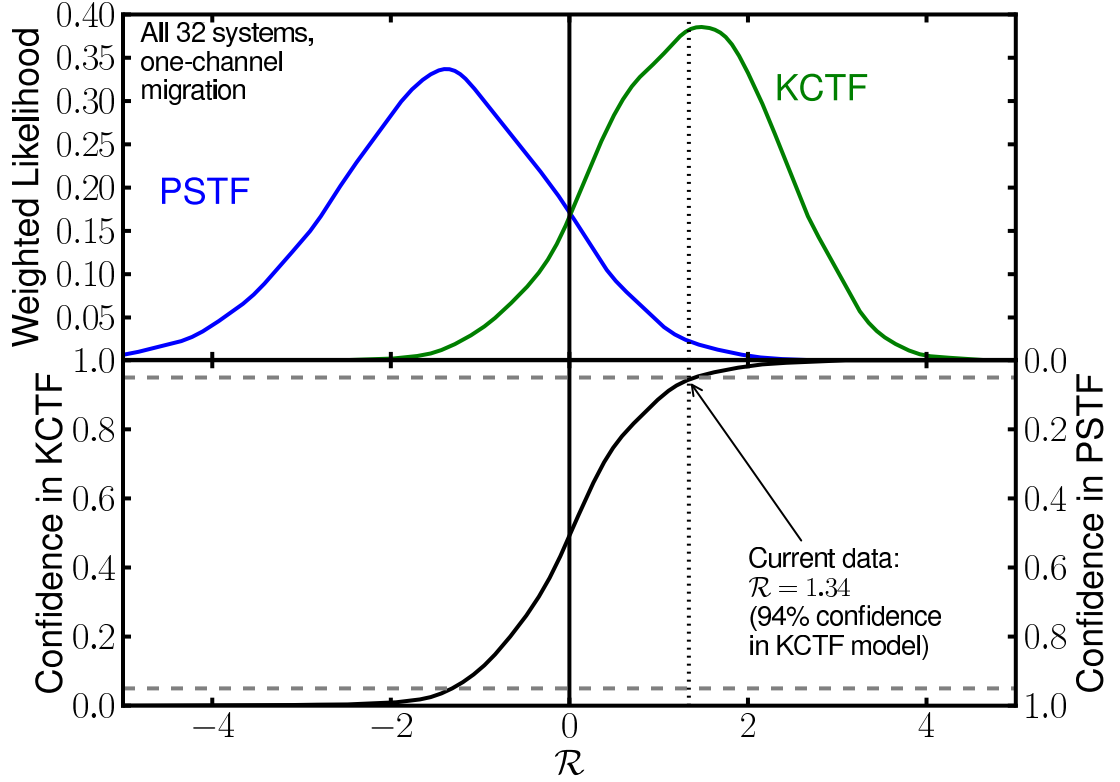


Figure 4.2 The relationship between our model selection statistic  $\mathcal{R}$ , or the logarithm of the likelihood ratio, and a model selection confidence. We randomly draw 5000 sets of 32  $\lambda$  values from each model and measure  $\mathcal{R}$  for each of these data sets. The top panel shows the distribution of  $\mathcal{R}$  values attained from these data simulations. The confidence in the model selection at any particular  $\mathcal{R}$  (bottom panel) is determined by the relative heights of the two distributions at that  $\mathcal{R}$ . The current data strongly favor the KCTF model.

data points to the FT07 KCTF model as explaining the formation of hot Jupiters better than the N08 model. However, as there are reasons to be skeptical that the Kozai mechanism could plausibly be responsible for the formation of all hot Jupiters, both theoretical (Wu et al. 2007) and empirical (Schlaufman 2010), we take our analysis one step further and consider what we may conclude if we allow for two distinct populations of systems: aligned and misaligned.

#### 4.4 Multiple migration channels?

Of the 32  $\lambda$  measurements to date, 14 have  $|\lambda| \leq 10$  and 19 are within  $2\sigma$  of  $\lambda = 0$ . Given that KCTF predicts many more systems with small  $\lambda$  than does the PSTF model, it is thus not surprising that it is preferred over PSTF in the above analysis. But what if only a fraction of planets migrate via a misaligning mechanism while the rest migrate through a process such as disk migration (Lin et al. 1996) that preserves spin-orbit alignment? Analyzing the first 10  $\lambda$  measurements (that included only a single significantly misaligned system), FW09

found such a two-population model to be their best selection. Especially given the difficulties of explaining all hot Jupiter migration using the Kozai mechanism alone (Wu et al. 2007), and given that Schlaufman (2010) found that fewer systems seem to be misaligned than the KCTF model predicts, it seems prudent to investigate how allowing for an aligned population affects conclusions about the intrinsic  $\psi$  distribution.

Accordingly, we add a parameter to each of our models:  $f$ , the fraction of systems that migrate according to either KCTF or PSTF, leaving the remaining  $1 - f$  fraction with  $\psi = 0$  (and thus  $\lambda = 0$ ). This requires us to modify our model selection procedure (§4.4.2), but it also enables us to ask an additional question: what do the models imply about  $f$  (§4.4.1)?

#### 4.4.1 What fraction of systems are misaligned?

We may use Bayes' theorem to calculate the probability distribution for the misaligned fraction  $f$  for each of our two models, conditioned on the 32 observed  $\lambda$  values. For this particular case, we may write Bayes' theorem as follows:

$$p_{f,\mathcal{M}}(f|\{\lambda\}) = \frac{\mathcal{L}_{\mathcal{M}}(\{\lambda\}|f)p(f)}{\int_0^1 \mathcal{L}_{\mathcal{M}}(\{\lambda\}|f)p(f)df} \quad (4.7)$$

where as before  $\{\lambda\}$  is the set of observed  $\lambda$  values and  $\mathcal{M}$  represents a particular model (either 'Koz' or 'scat'),  $p_{f,\mathcal{M}}(f|\{\lambda\})$  is the posterior probability distribution for  $f$  under model  $\mathcal{M}$ ,  $\mathcal{L}_{\mathcal{M}}(\{\lambda\}|f)$  is the likelihood of the data given a particular  $f$  under  $\mathcal{M}$ , and  $p(f)$  is the prior probability distribution for  $f$ , which we take to be flat between 0 and 1. The denominator is the normalization factor, also known as the marginal likelihood. The posterior probability distribution for  $f$  allows us to infer conclusions about  $f$  for a particular model given the current data. The likelihood function is calculated the same way as Eq. 4.2, except that now the probability distribution for  $\lambda$  is dependent on  $f$ :

$$p_{\lambda,\mathcal{M}}(\lambda|f) = f \times p_{\mathcal{M}}(\lambda) + (1 - f)\delta(\lambda), \quad (4.8)$$

where  $p_{\mathcal{M}}(\lambda)$  are the  $\lambda$  distributions we calculated in §4.2 (bottom panel of Figure 4.1), and  $\delta(\lambda)$  is the Dirac delta function.

Fig. 4.3 illustrates  $\mathcal{L}_{\mathcal{M}}(\{\lambda\}|f)$  as a function of  $f$  for each model. We measure the most likely values and their symmetric 95% confidence ranges for  $f$  by normalizing the likelihoods and computing the cumulative distribution functions. If the KCTF + aligned model is correct,  $f$  lies between 0.40 and 0.87 with 95% confidence, with the most likely value being 0.64. Similarly, if the PSTF + aligned model is correct,  $f$  lies between 0.34 and 0.76 with 95% confidence, with the most likely value being 0.55. Both models indicate a significant fraction of aligned systems, with the PSTF model indicating that nearly half of systems might be intrinsically aligned. So the next question to ask is in this two-channel picture of planet migration, which model do the data favor?

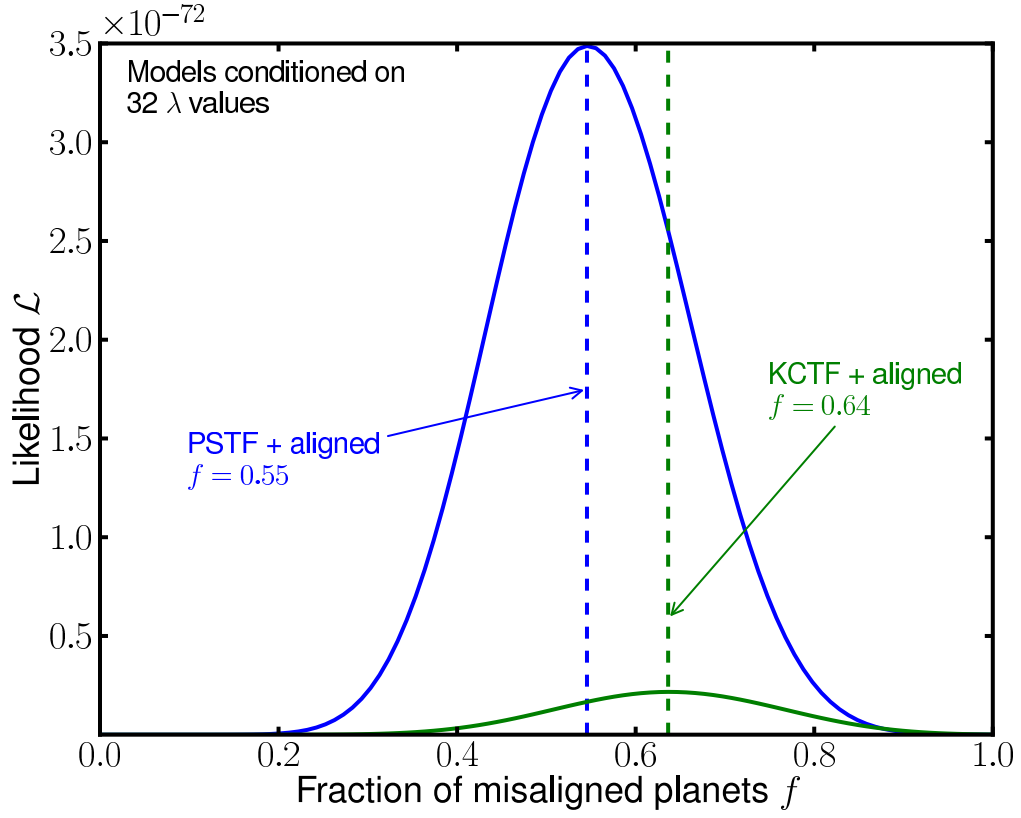


Figure 4.3 The posterior distribution of  $f$  for each of our two-migration-channel models, conditioned on 32 observed  $\lambda$  values. We show that if the true misaligned population were misaligned according to the KCTF distribution, then we expect that about 64% of close-in planets are misaligned. Similarly, if the misaligning mechanism were PSTF, then the most likely value of  $f$  is close to 55%. Strictly speaking, this figure shows the likelihood of the current data as a function of  $f$  for each model; if these curves were normalized they would be proper probability distributions, given our flat prior for  $f$ . However, plotting them in un-normalized form is illustrative, since we use the ratio of the areas under these curves (the marginal likelihood) as our two-mode model selection criterion.

#### 4.4.2 Two-mode model selection and confidence assessment

Since the models we are now comparing each have an unknown parameter  $f$ , we redefine our model comparison statistic to be the ratio of the marginal likelihoods (the denominator of Eqn. 4.7, or the area under the curves in Figure 4.3) of the two models:

$$\mathcal{R} = \log_{10} \left( \frac{\int_0^1 \mathcal{L}_{\text{Koz}}(\{\lambda\}|f)p(f)df}{\int_0^1 \mathcal{L}_{\text{scat}}(\{\lambda\}|f)p(f)df} \right), \quad (4.9)$$

where again  $\mathcal{R} > 0$  favors KCTF + aligned and  $\mathcal{R} < 0$  favors PSTF + aligned. This time, we calculate  $\mathcal{R} = -1.16$ , which favors PSTF over KCTF.

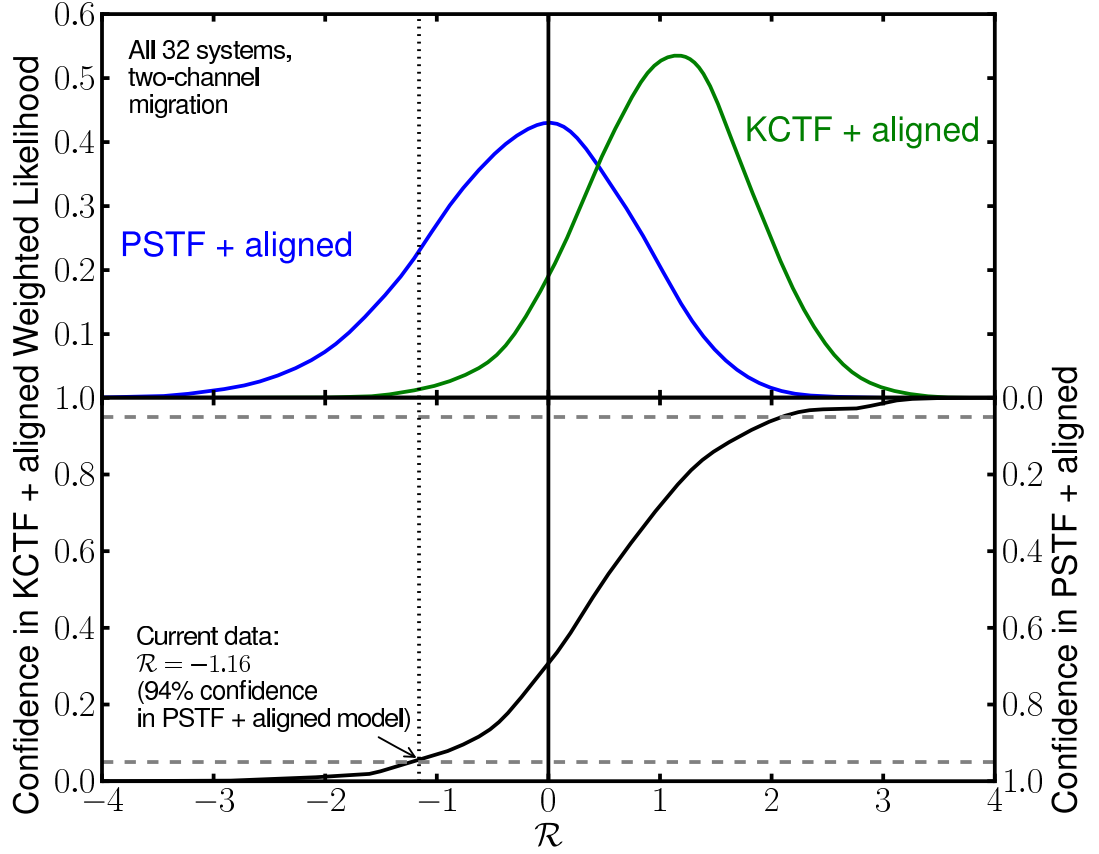


Figure 4.4 The relationship between our model selection statistic and model selection confidence for comparing our two-mode migration models, where only a fraction  $f$  of planets are misaligned and the rest are in perfectly-aligned systems.  $\mathcal{R}$  in this case is the logarithm of the ratio of the marginal likelihoods of the two models, and the  $\mathcal{R}$  distributions are generated by measuring the  $\mathcal{R}$  values for data simulations at different  $f$ . The PSTF + aligned model is preferred over the KCTF + aligned model. See §4.4.2 for details.

Confidence assessment requires data simulations as before, except now we simulate data sets for each model on a grid of  $f$  values, to determine the behavior of  $\mathcal{R}$  as a function both of misalignment model and  $f$ . So for each of 20 equally-spaced values of  $f$  between 0 and 1, we randomly draw 1000 sets of 30 systems, as described in §4.3.2, simulating the aligned fraction of planets by giving each simulated  $\lambda$  a probability  $(1-f)$  to be re-assigned to  $\lambda = 0$  before being perturbed by the measurement error.

These simulated  $\mathcal{R}$  values give us an  $\mathcal{R}$  PDF for each  $f$  for each model—essentially empirical two-dimensional likelihood functions:  $\mathcal{L}_{\mathcal{R},\text{Koz}}(\mathcal{R}, f)$  and  $\mathcal{L}_{\mathcal{R},\text{scat}}(\mathcal{R}, f)$ , where  $\mathcal{R}$  is continuous but  $f$  is only sampled at 20 points between 0 and 1. Since we already calculated the posterior probability distribution of  $f$  given the current data for each of our models above, we may marginalize these likelihood functions over  $f$  to

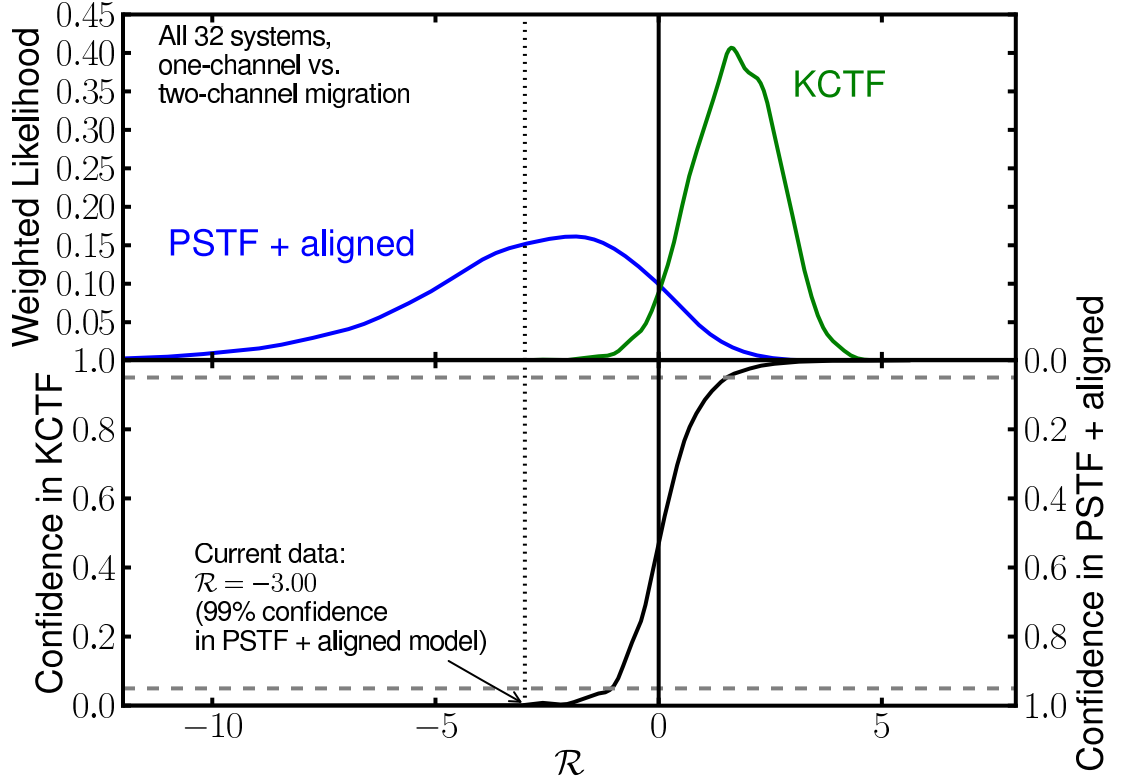


Figure 4.5 The relationship between our model selection statistic and model selection confidence, comparing the preferred single-mode migration scenario (KCTF) to the preferred two-mode model (PSTF + aligned). PSTF + aligned is strongly preferred over single-mode KCTF.

calculate a properly weighted one-dimensional PDF for  $\mathcal{R}$ :

$$p_{\mathcal{R},\mathcal{M}}(\mathcal{R}) = \sum_{i=1}^{20} \mathcal{L}_{\mathcal{R},\mathcal{M}}(\mathcal{R}, f_i) \Pr(f_i | \{\lambda\}, \mathcal{M}). \quad (4.10)$$

The term  $\Pr(f_i | \{\lambda\}, \mathcal{M})$  is the probability obtained by integrating  $p_{f_i,\mathcal{M}}(f | \{\lambda\})$  (Eqn. 4.7; Figure 4.3) between  $f_i - \Delta f/2$  and  $f_i + \Delta f/2$ , where  $\Delta f$  is the spacing between successive  $f$  values in our simulations. Using these likelihood functions, we may calculate the probability of either of our models being true, analogously to Eqns 4.5 and 4.6. We find that our measurement of  $\mathcal{R} = -1.16$  gives a confidence of 94% for the PSTF model, as illustrated in Figure 4.4.

Thus, while the current sample of  $\lambda$  measurements appears at first to be evidence for Kozai migration as predicted by FT07, the model preference changes in favor of the N08 scattering model if allowance is made for the existence of some systems forming hot Jupiters via some mechanism that preserves alignment. And since what current RM measurements can tell us about planet migration mechanisms changes significantly depending on whether we assume one or two migration channels, we next explore whether current data allows



us to distinguish between those scenarios.

#### 4.4.3 One channel or two channels?

We approach this question the same way we have hitherto approached the other model selection questions: define a model selection statistic  $\mathcal{R}$ , calculate  $\mathcal{R}$  for the current data, and determine confidence based on data simulations. Our model selection statistic in this case becomes:

$$\mathcal{R} = \log_{10} \left( \frac{\mathcal{L}_{\text{one}}}{\mathcal{L}_{\text{two}}} \right), \quad (4.11)$$

where  $\mathcal{L}_{\text{one}}$  is one-mode likelihood as used in Eq. 4.1, and  $\mathcal{L}_{\text{two}}$  is a marginalized likelihood as used in Eq. 4.9. Since KCTF is preferred for one-mode migration and PSTF is the preferred two-mode model, we take  $\mathcal{L}_{\text{one}}$  to be the likelihood of the data under the KCTF model and  $\mathcal{L}_{\text{two}}$  to be the marginal likelihood of the data under the PSTF+aligned model. The advantage of the Bayesian approach here is that it allows us to properly compare models with different numbers of parameters (in this case, zero and one). We calculate  $\mathcal{R} = -3.00$ , which strongly favors the PSTF + aligned model over the one-mode KCTF model, with a confidence of 99% (Figure 4.5).

### 4.5 Hot Stars

We have shown that the KCTF prediction for the distribution of  $\psi$  for migrated planets does not adequately explain the observed distribution of  $\lambda$ , and that current data favor a combination of well aligned systems and systems misaligned via planet-planet scattering. However, it may be that the current population we observe is not representative of the initial misalignment distribution. For example, Matsumura et al. (2010) conclude that tidal effects may be important in damping out initial misalignment in some systems. Winn et al. (2010) note that most of the misaligned planets that have been observed to date are around stars hotter than 6250 K. Based on this empirical finding, they suggest that perhaps all hot Jupiters (HJs) migrate through some misaligning mechanism, and that cooler stars with deep convective zones have their envelopes tidally torqued into alignment by the planet on a relatively quick timescale, thereby erasing the evidence of the initial misalignment. If this were indeed the case, then an important key to understanding HJ migration would be spin-orbit measurements of planets transiting hot stars, since these systems would presumably have maintained their primordial misalignment.

Following this line of inquiry, we repeat the analyses of §3 and §4 restricted to only the current sample of 12 hot stars ( $T_{\text{eff}} > 6250$  K). Of our two one-mode migration models, PSTF is favored over KCTF, with 82% confidence ( $\mathcal{R} = -0.73$ ; Figure 4.6). We also explore the two-mode migration hypothesis, allowing for hot stars to also have an intrinsically aligned population. The posterior distributions for  $f$  under the two models conditioned on the hot star data are illustrated in Figure 4.7, showing that both models favor almost complete

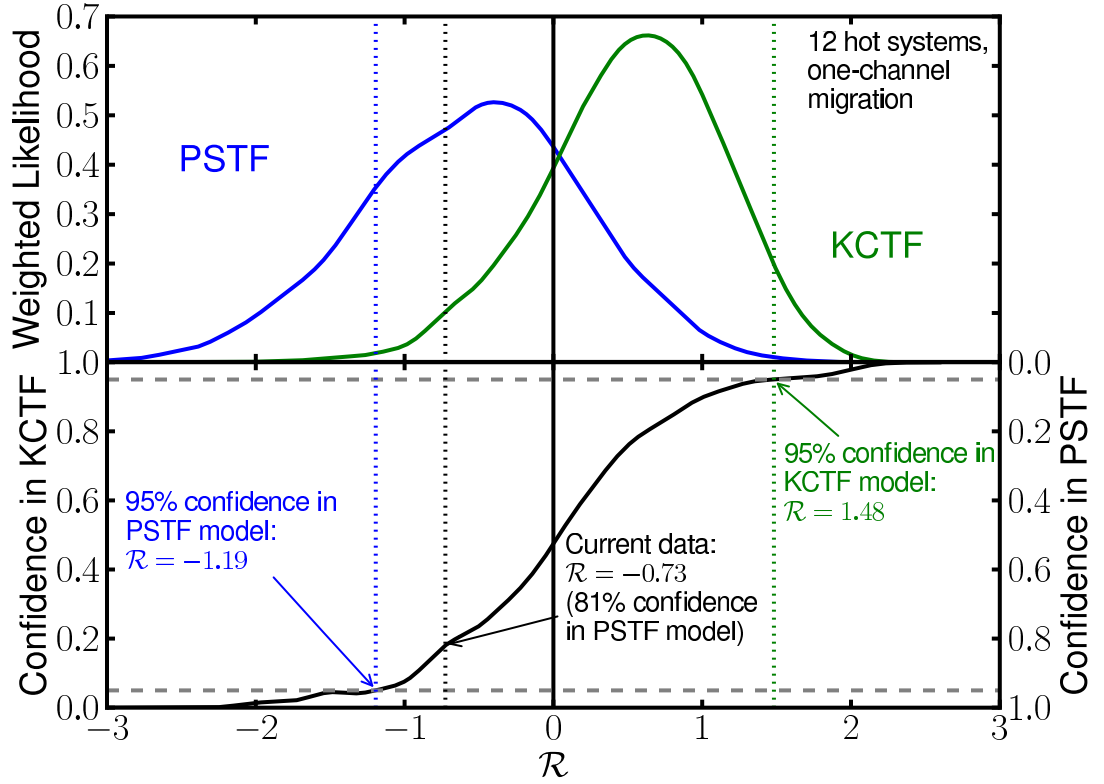


Figure 4.6 The relationship between our model selection statistic and model selection confidence, but using only the 12 hot stars ( $T_{\text{eff}} > 6250$ ) as our data. One-channel migration favors the PSTF mechanism for this subset of the data, but not conclusively. The  $\mathcal{R}$  thresholds required to reach 95% confidence in either model for this sample size are marked.

misalignment. In fact, one-mode vs. two-mode model selection for this subset of the data gives  $\mathcal{R} = -0.33$ , favoring the single-mode hypothesis. For this reason we do not pursue any further the idea that there exists an intrinsically aligned population among hot stars, as the current data do not merit the additional model complication.

Thus, if the distribution of hot star  $\lambda$  values represents the primordial alignment distribution for all stars, the current 12  $\lambda$  observations hint that the PSTF prediction of N08 describes planet migration better than the KCTF model of FT07. However, there are not yet enough data for this model selection to be conclusive. In the next section, we discuss how many more RM measurements will likely be needed in order to make hot star model selections confident.

## 4.6 How many hot star RM observations are needed?

With the full sample of 32 RM observations to date, we are able to confidently state that a combination of well aligned systems and systems misaligned via planet-planet scattering (the PSTF model) explains the

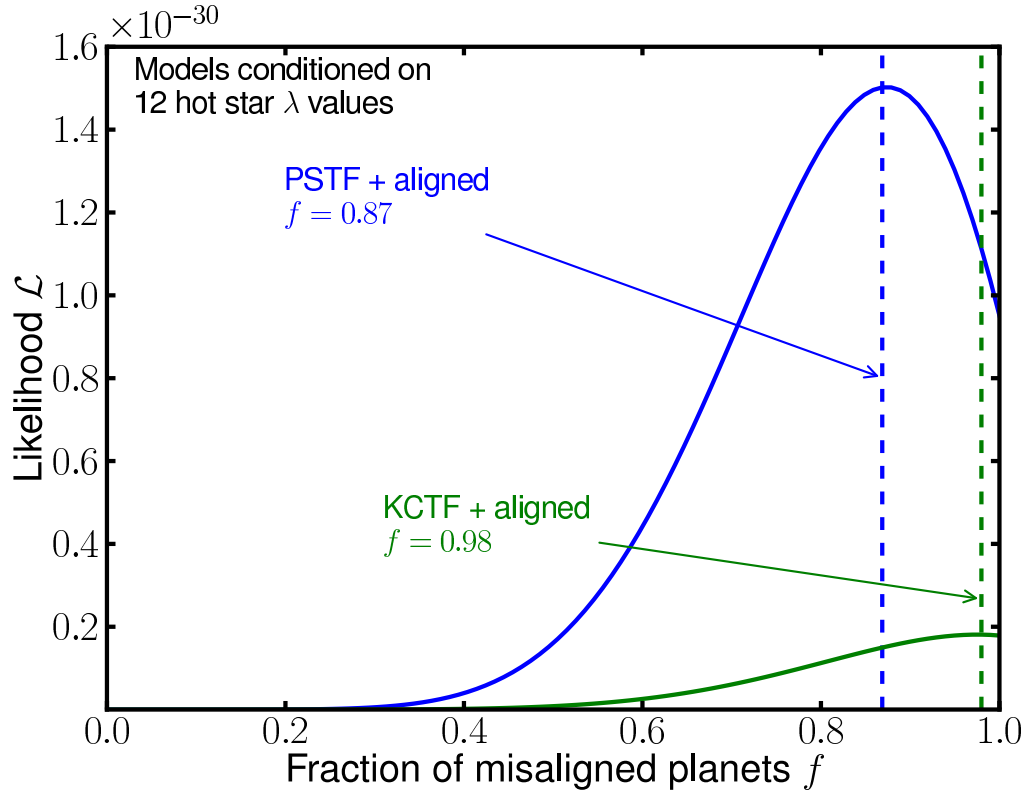


Figure 4.7 The posterior distribution of  $f$  for each of our two-migration-channel models, conditioned on the 12 observed hot star ( $T_{\text{eff}} > 6250$ )  $\lambda$  values. As Winn et al. (2010) point out, hot stars appear to be misaligned more often than the overall stellar population. Because  $f = 1$  has a high probability under both models, we do not further pursue two-mode migration model comparison for this subsample.

current data better than migration via the Kozai mechanism and tidal friction alone (KCTF). Only considering the sample of 12 hot stars, our present conclusions are weaker. Inspired by the work of Swift & Beaumont (2010), who calculated the sample size of protostellar cores necessary to reliably distinguish a power law from a lognormal distribution, we wish to quantify how many additional hot star  $\lambda$  measurements will likely be needed in order to measure  $\mathcal{R}$  values indicative of confident ( $> 95\%$ ) model selection between the KCTF and PSTF models.

This requires determining two things: first, what values of  $\mathcal{R}$  correspond to 95% confident model selections for a given sample size  $N$ , and second, how likely we are to actually measure such a confident value once we have  $N$  observations in hand, given the current set of 12 as our starting point.

Both questions can be addressed by data simulations. The  $\mathcal{R}$  thresholds that represent 95% confidence for a given  $N$  can be determined by simulating many data sets of sample size  $N$  and using Eq. 4.5 to calculate model selection confidence as a function of  $\mathcal{R}$ . This has already been illustrated for  $N = 32$  in Figures 4.2, 4.4, and 4.5 and for  $N = 12$  in Figure 4.6. We repeat this procedure for different sample sizes, defining  $\mathcal{R}$  thresholds for our model selection for a series of  $N$  values up to  $N = 100$ .

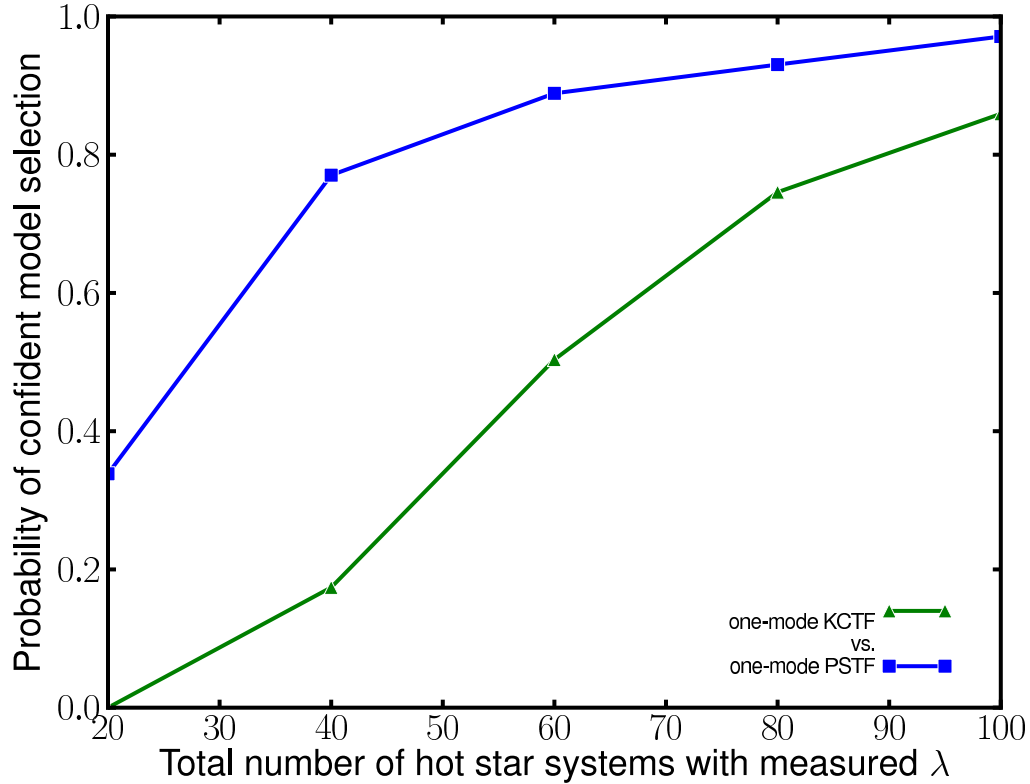


Figure 4.8 The probability with which a given-sized future sample of hot-star ( $T_{\text{eff}} > 6250$  K)  $\lambda$  values (in addition to the 12 that currently exist) will result in a confident selection ( $> 95\%$ ) of one of our misalignment models over another. We use data simulations of various sample sizes (described in §4.6) to define the confidence thresholds and to determine how likely those thresholds are to be reached for each particular sample size. For example, with 28 more measurements (40 total), we have a 75% chance of selecting the PSTF model at  $> 95\%$  confidence.

We also use data simulations to determine the probability of a future experiment resulting in a confident  $\mathcal{R}$  measurement. To do this we again simulate many data sets of the same sizes as above and measure the  $\mathcal{R}$  distributions of the experiments, but this time the first 12 measurements in each data set are fixed to be the current hot star measurements. Then we use the  $\mathcal{R}$  distributions to predict how often threshold values are reached for each model. This probability is plotted against sample size in Figure 4.8.

We learn from this that in order to have  $> 90\%$  chance of obtaining a hot star data set that confidently selects either of our single-mode migration models, we will likely need a total data set of about 80-100 hot star RM observations (Figure 4.8). More optimistically, there is  $> 50\%$  chance of confident model selection with only a factor of  $\sim 3$  more observations ( $\sim 35$  total) if the PSTF model is correct. Given the pace at which this field is growing, this may be reasonably expected to happen within the next few years. On the other hand, if the KCTF model (or some other model) is a better description of reality, then it will likely take more observations to determine this, given the preference of the current data for the PSTF model.

Table 4.1. Results summary

Data	Model 1	Model 2	$f^a(\text{model 1})$	$f^a(\text{model 2})$	$\mathcal{R}^b$	Confidence <sup>c</sup>
All 32 systems	<b>KCTF<sup>d</sup></b>	PSTF <sup>e</sup>	—	—	1.34	96%
All 32 systems	KCTF + aligned	<b>PSTF + aligned</b>	$0.64 \pm 0.24$	$0.55 \pm 0.21$	-1.16	93%
All 32 systems	KCTF	<b>PSTF + aligned</b>	—	$0.55 \pm 0.21$	-3.00	99%
12 hot systems <sup>f</sup>	KCTF	<b>PSTF</b>	—	—	-0.73	81%

— indicates favored model.

<sup>a</sup>maximum-likelihood value, with symmetric 95% confidence range; only applicable to two-mode models

<sup>b</sup> $\mathcal{R} > 0$  favors model 1;  $\mathcal{R} < 0$  favors model 2

<sup>c</sup>Degree of belief that the model selection is correct; based on Monte Carlo simulations (§3.2)

<sup>d</sup>Kozai cycles + tidal friction (Fabrycky & Tremaine 2007)

<sup>e</sup>Planet-planet scattering + Kozai cycles and tidal friction (Nagasawa et al. 2008)

<sup>f</sup>host star  $T_{\text{eff}} > 6250$  K

. Bold text

## 4.7 Discussion and Conclusions

The study of exoplanet spin-orbit angles is advancing rapidly, with 32 measured projected spin-orbit angles and more to come as ground- and space-based surveys continue to detect transiting planets. We have analyzed the current sample and quantify what may be inferred about the distribution of true spin-orbit angles  $\psi$ . In particular, we ask whether the current data are sufficient to test the predicted distribution of  $\psi$  from specific migration mechanisms, using the Kozai cycles + tidal friction (KCTF) model of Fabrycky & Tremaine (2007) and the planet-planet scattering model of Nagasawa et al. (2008) (that also includes the Kozai effect and tidal friction; PSTF) as test models.

We find that conclusions about which migration mechanism is responsible for misalignment depend crucially on the assumption of whether there exists a population of intrinsically aligned systems ( $\psi = 0$ ). Without allowing for an intrinsically aligned population we find that the KCTF model is favored over the PSTF mechanism (§4.3), but allowing for this population we find that PSTF is favored (§4.4), with the most likely fraction of misaligned systems being 0.55, and between 0.34 and 0.76 with 95% confidence. We also find that this two-mode migration model (PSTF + aligned) is significantly favored over single-mode KCTF migration. This agrees with Schlaufman (2010), who also concluded that there is likely to be an aligned population, based on detecting fewer likely-misaligned systems than expected according to the predictions of the KCTF model.

These results may be an indication of two migration channels for close-in planets, one that acts gently, preserving the alignment of planet orbits with the stellar spin, and one that acts impulsively, causing misalignment. The gentle mechanism might well be disk migration (Lin et al. 1996), and our analysis suggests that the misaligning mechanism is not solely the Kozai effect but rather some mechanism that distributes  $\psi$  more broadly, such as planet-planet scattering in combination with the Kozai effect. This accords with the

conclusion of Matsumura et al. (2010) that some of the close-in planets with non-zero orbital eccentricities are likely to have been formed by planet-planet scattering and subsequent tidal circularization.

Focusing on the subsample of 12 hot star ( $T_{\text{eff}} > 6250$  K)  $\lambda$  measurements, which Winn et al. (2010) predict might be the only systems to maintain their primordial misalignments, we find that the data prefer the single-mode PSTF model over KCTF, with a confidence of 85%. We also find that a single migration mechanism is sufficient to describe the current hot star  $\lambda$  distribution, without including an intrinsically well aligned fraction.

Looking to the future, we also calculate the number of additional hot star  $\lambda$  measurements necessary to achieve  $> 95\%$  confidence in the hot star model selection (§4.6). We find that if either of our single-mode mechanisms describes the  $\psi$  distribution around hot stars, then a total data set of about 80-100  $\lambda$  measurements should definitely be sufficient to solidify which is the preferred model, with a  $> 50\%$  probability of confident model selection with a total data set of only about 40, if scattering is indeed the best explanation for the  $\psi$  distribution of close-in planets.

Thus, we suggest that if RM studies wish to be able to identify migration mechanisms through measuring  $\lambda$  distributions, they should concentrate on measuring  $\lambda$  for planets around hot stars. Of course it is conceivable that migration mechanisms themselves might be different around different types of stars, in which case hot star  $\lambda$  measurements might not tell us anything about cool star migration, but that is the kind of question that will only be able to be explored when much more data is available.

Finally, we emphasize that the results in this paper are illustrative more than definitive, as we have only tested two very specific misalignment models. Consequently, we encourage continued theoretical work predicting  $\psi$  distributions, as the analysis we present may be applied to any prediction. We can use the results of this paper as a guide for what to expect from such future analyses. For example, models that favor larger values of  $\lambda$  more strongly than the present KCTF prediction (as does the PSTF model) are likely to be preferred. We also suggest that an interesting question to pursue would be self-consistent planet formation and dynamical evolution calculations, in order to explore planet-planet scattering in the context of realistic formation scenarios (in contrast to the fixed initial conditions of the Nagasawa et al. (2008) simulations). A particularly intriguing angle to explore in such work would be whether a plausible explanation for the observed trend in  $\lambda$  with stellar temperature/mass might be explainable by more massive stars tending to form more massive planets in closer proximity, so as to make planet-planet scattering (and thus spin-orbit misalignment) more common among earlier-type stars. Migration models might also be combined with models that predict that large values of  $\psi$  might originate from the host star itself being tilted relative to the disk (Lai et al. 2010), rather than solely from the effects of planet migration, though current observations suggest that protoplanetary disks tend to be aligned with stellar spins (Watson et al. 2010).

Planet migration has been a mystery ever since the first hot Jupiters were discovered, with very little observational evidence to substantiate theories. As fossil remnants of planet migration, spin-orbit angles are key to understanding the origins of these close-in planets, and the first few years' worth of  $\lambda$  measurements

are beginning to give substantial clues. Many more transiting planets will be discovered in the near future thanks to the productivity of transit surveys, and as long as RM measurements of these systems continue, our understanding of planet migration will continue to improve.

## Chapter 5

# On the Low False Positive Probabilities of Kepler Planet Candidates

*This chapter is adapted from “On the Low False Positive Probabilities of Kepler Planet Candidates,” ApJ 738, 170 (2011), by TDM and John Johnson. This paper was first posted nearly simultaneous with the release of the first public Kepler planet candidate list (Borucki et al. 2011). The inspiration for this work was some quick back-of-the-envelope calculations that convinced me that the false positive rate of well-vetted Kepler candidates should be significantly lower than that of ground-based transit surveys. As the Kepler team did not present anything more than very rough qualitative comments regarding the false positive rate of the candidate sample, the conclusions presented here continue to be the basis for many statistical studies that make conclusions about exoplanets based on the Kepler candidate lists.*

### 5.1 Introduction

In the wake of the first full release of planet candidates from the *Kepler* mission (Koch et al. 1998; Borucki et al. 2008, 2011), the study of the properties of exoplanetary systems has entered a new era. For the first time there exists a large uniform sample of transiting planets largely unaffected by the detection challenges and selection effects inherent in ground-based searches (Gaudi 2005; Gaudi et al. 2005, e.g.), enabling the first clear glimpse of the population of exoplanets down to the size of Earth as well as the first opportunity to study planet radii at large orbital separations. However, follow-up observations to unambiguously confirm individual signals are time-consuming and difficult (or impossible), especially for fainter stars and smaller planets. Consequently, in order to understand what the population of *Kepler* transit-like signals can tell us about the population of exoplanets in general, the problem of astrophysical false positives must be understood.

From the early days of planet transit searches, eclipsing binary systems masquerading as transit signals have plagued detection efforts (Konacki et al. 2003; O’Donovan et al. 2006; Poleski et al. 2010; Almenara et al. 2009). Generally speaking, there are three types of astrophysical false positive: a grazing eclipsing binary, a dwarf star eclipsing a giant star, and a blended eclipsing binary system, which may be either a



hierarchical triple system or an unassociated binary blended within the aperture of a target star (Torres et al. 2004).<sup>1</sup>

The remarkable photometric precision that *Kepler* is delivering (Jenkins et al. 2010b,  $\sim 30$  ppm) allows for an immediate simplification of the false positive landscape. Batalha et al. (2010b) explain the multitude of ways that certain common false positive scenarios can be identified from *Kepler* photometry alone. For example, grazing eclipsing binaries can be identified by their V-shaped transits, and the giant-eclipsed-by-a-dwarf scenario can be avoided both by the comprehensive work that went into assembling the Kepler Input Catalog (Latham et al. 2005; Batalha et al. 2010a) and by the ability to photometrically identify giants by their elevated levels of stellar variability compared with dwarf stars (Basri et al. 2010). Even many blended binaries can be identified from the *Kepler* photometry and astrometry alone, by looking for a shift in the center of light, e.g the “rain diagrams” of Jenkins et al. (2010a). However, some blended binary scenarios remain undetectable by this technique, especially those in hierarchical triple systems, and so a detailed understanding of the false positive problem for *Kepler* requires a detailed understanding of the probability of encountering such blend scenarios.

The *Kepler* team has proven that extremely careful and detailed analyses of individual systems can “validate” planets probabilistically by combining various follow-up observations with modeling the light curves of all possible false positive scenarios with the so-called BLENDER software (Torres et al. 2011). However, this method is computationally expensive and labor-intensive, rendering it a time-consuming process. As a result, only three BLENDER-validated planets having been revealed to date: Kepler-9d (Torres et al. 2011), Kepler-11g (Lissauer et al. 2011), and Kepler-10c (Fressin et al. 2011). With dedicated supercomputer resources coming online for the *Kepler* team’s use, this number will certainly rise, but the fact remains that it will be a long time before the BLENDER method can be applied to any large number of the *Kepler* candidates (*Kepler* team, 2011, private comm.); in the meantime statistical interpretations of the candidate sample will rely on statistical assumptions of the false positive rate.

There has been significant previous effort in the literature dedicated to predicting the expected rate of false positive transit signals. Brown (2003) pioneered this work by predicting the rates of different types of false positives and Jovian planet detections for a variety of different surveys, including the then-future *Kepler* mission. Evans & Sackett (2010) greatly extend this work by deriving detection and false positive rates from full-scale bottom-up simulations of synthetic ground-based transit surveys, taking into account all false positive possibilities and many details not included by Brown (2003). We continue in the tradition of these authors with an analysis directly applicable to the *Kepler* mission, approaching from a slightly different angle. Instead of focusing on predicting an overall number or expected rate of planet detections or false positives, we instead seek a simple answer to the following question: “What is a conservative estimate of the probability that an observed apparent transit signal is in fact a true transiting planet?” By framing the issue in

---

<sup>1</sup>In this paper we consider “false positives” to be purely stellar configurations mimicking transiting planet signals. For discussion of scenarios involving “blended planets,” see Appendix.

this manner we are able to sidestep the complex issue of detectability, as our analysis assumes a transit-like signal has been detected.

Our philosophy in this work is not to take into account all conceivable details of transit signals, but rather to consider only those which are most salient: the brightness of the *Kepler* target star, its location in the field, and transit signal depth. The details we choose not to address in this work (notably transit period and duration) are those we judge would add uncertainty to our calculations while tending to only decrease our estimates of the false positive probability. Thus we are able to keep our analysis straightforward, yet remain confident we are calculating conservative upper limits to the probability that any given *Kepler* transit signal might be a false positive. As we show in §5.2 and again in §5.3, even these conservative upper limits are enough to indicate that *Kepler* planet candidates will only rarely turn out to be false positives.

## 5.2 Basic Bayesian Framework

The probability that a given transit signal is of planetary origin may be expressed as the following, according to Bayes' theorem:

$$\Pr(\text{planet} \mid \text{signal}) = \frac{\Pr(\text{signal} \mid \text{planet})\Pr(\text{planet})}{\Pr(\text{signal})}. \quad (5.1)$$

In this framework  $\Pr(\text{signal} \mid \text{planet})$  is the probability of obtaining the observed signal given that there is a transiting planet on an orbit of a particular period. This factor is known as the *likelihood* of the signal under the planet hypothesis, and we will abbreviate it as  $\mathcal{L}_{\text{pl}}$ .  $\Pr(\text{planet})$  is the probability of a star hosting a transiting planet (the occurrence rate of planets times the transit probability), which must enter the calculation as an *a priori* assumption. Thus we call this factor, according to Bayesian convention, the *prior* on planets, and designate it  $\pi_{\text{pl}}$ .

Since there are only two possible origins of a transit-like signal (planet or false positive), the denominator of Equation 5.1 can be rewritten as marginalizing over the possible models:

$$\Pr(\text{signal}) = \mathcal{L}_{\text{pl}}\pi_{\text{pl}} + \mathcal{L}_{\text{FP}}\pi_{\text{FP}}. \quad (5.2)$$

Using our convention,  $\mathcal{L}_{\text{FP}}$  and  $\pi_{\text{FP}}$  are the likelihoods and priors for a false positive signal. The false positive term can be further broken down accounting for the two specific false positive scenarios we are exploring: the blended eclipsing binary (BB) and the hierarchical eclipsing triple (HT), allowing Equation 5.1 to be rewritten as the following:

$$\Pr(\text{planet} \mid \text{signal}) = \frac{\mathcal{L}_{\text{pl}}\pi_{\text{pl}}}{\mathcal{L}_{\text{pl}}\pi_{\text{pl}} + \mathcal{L}_{\text{BB}}\pi_{\text{BB}} + \mathcal{L}_{\text{HT}}\pi_{\text{HT}}}. \quad (5.3)$$

In general, the likelihoods depend on the particularities of the transit signal and enable discrimination between models depending on the transit depth, shape, or period. For now we ignore these details, assuming

for the moment that we have no knowledge of the differences between the kind of transit signals to expect from planets and from false positives. This enables us to write a simplified version of Eq. 6.2:

$$\Pr(\text{planet} \mid \text{signal}) \approx \frac{\pi_{\text{pl}}}{\pi_{\text{pl}} + \pi_{\text{BB}} + \pi_{\text{HT}}}. \quad (5.4)$$

We then define the “false positive probability” (FPP) as the complement of this probability:

$$\text{FPP} = 1 - \Pr(\text{planet} \mid \text{signal}) \quad (5.5)$$

Thus, before considering any detailed information of a particular light curve, the probability that an observed transit signal is actually a false positive depends only on the relative occurrence rates of planets and the false positive scenarios. As mentioned above,  $\pi_{\text{pl}}$  is simply an assumed occurrence rate of planets times the transit probability; we explain how we determine  $\pi_{\text{BB}}$  and  $\pi_{\text{HT}}$  in the following subsections. We describe first this priors-only framework in order to elucidate what dominates our final results, but in §5.3 we will include the likelihoods we removed in Equation 5.4, taking into account dependence on the depth of the transit signal.

### 5.2.1 Blended Binaries

The probability of a transit-mimicking binary system to be blended within the aperture of a *Kepler* target star ( $\pi_{\text{BB}}$ ) can be broken down in the following way:

$$\pi_{\text{BB}} = \Pr(\text{blend}) \cdot \Pr(\text{appropriate eclipsing binary}). \quad (5.6)$$

The first factor is the probability for a potentially blending star to be projected within a given radius of a *Kepler* star, and the second is the probability for that star to be an eclipsing binary system that can appropriately mimic a planetary transit.

To calculate these probabilities, we use the stellar population synthesis and Galactic structure code TRILEGAL (TRIdimensional model of the GALaxy; Girardi et al. (2005)), which is publicly available on the web<sup>2</sup>. TRILEGAL simulates the physical and photometric properties of the stars along a given line of sight, using various stellar evolution grids (Girardi et al. 2002; Chabrier et al. 2000) and a Galactic model that includes a halo, thin and thick disks, and a bulge. All of our simulations use a Chabrier lognormal IMF (Chabrier 2001) and default TRILEGAL values for the Galactic structure parameters, including a squared hyperbolic secant structure for the thin disk, an exponential structure for the thick disk, and an oblate spheroid for the halo.

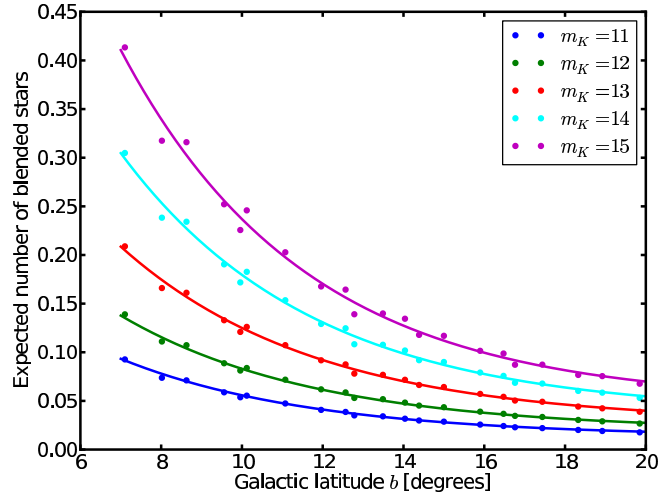


Figure 5.1 The probability for a possibly blending star to be projected within  $2''$  of a *Kepler* target star, as a function of Galactic latitude, as determined by TRILEGAL simulations. The plotted points are simulations; the lines are the exponential fits as described in Equation 5.8.

#### 5.2.1.1 Probability of a blend

The blend probability can be calculated by determining the average sky density (e.g. stars per square arcsec) of stars faint enough so as not to be obviously present in *Kepler* data yet bright enough to possibly mimic a transit. The first condition is somewhat subjective, and we conservatively say that a star must be more than 1 magnitude fainter than the *Kepler* primary in order to be able to hide undetected within the *Kepler* aperture. In practice the true value is probably significantly fainter, but this approximation will lead to only a small overestimate of the blended star probability, as there are many more faint than bright stars.

The faint condition can be determined by noting that in order for a blended eclipsing binary system to mimic a transit of fractional depth  $\delta$ , the blended system must comprise more than a fraction  $\delta$  of the total flux within the *Kepler* aperture. This condition may be expressed as the following:

$$m_{K,\text{bin}} - m_{K,\text{target}} = \Delta m_K \leq -2.5 \log_{10}(\delta), \quad (5.7)$$

where  $m_{K,\text{bin}}$  is the total apparent *Kepler* magnitude of the blended binary system and  $m_{K,\text{target}}$  is the magnitude of the *Kepler* target star. A transit depth of  $\delta = 0.01$  corresponds to  $\Delta m_K = 5$ ; for  $\delta = 10^{-3}$ ,  $\Delta m_K = 7.25$ ; and for  $\delta = 10^{-4}$  (approximately an Earth-sized transit of a Solar-radius star),  $\Delta m_K = 10$ . This means that no binary system fainter than  $m_K = 24$  can possibly mimic a  $\delta = 10^{-4}$  transit around a  $m_K = 14$  star, which is a typical magnitude for a *Kepler* target.

Using TRILEGAL, we determine the sky density of stars in this magnitude range within in the *Kepler* field, and thus the probability of one by chance being projected close to a *Kepler* target star, by simu-

<sup>2</sup><http://stev.oapd.inaf.it/cgi-bin/trilegal>

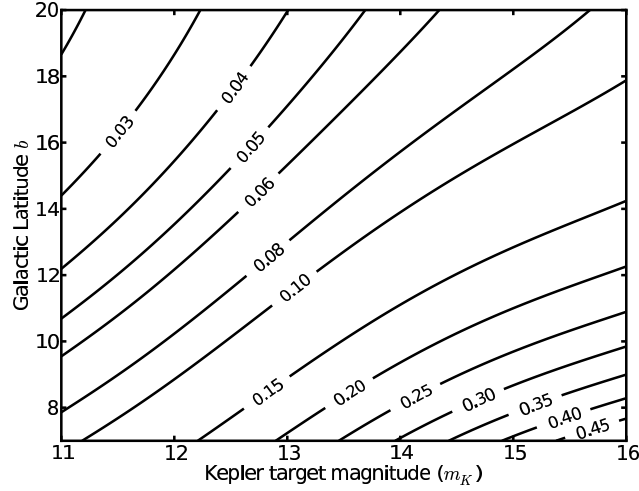


Figure 5.2 The probability for a possibly blending star to be projected within  $2''$  of a *Kepler* target star, as a function of both Galactic latitude and target star magnitude, as determined by TRILEGAL simulations.

lating a  $10 \text{ deg}^2$  field centered on the center of the *Kepler* field. We then simply count the stars within the desired range of Kepler magnitude (which TRILEGAL provides). As a fiducial example, the average density of stars between  $m_K = 15$  and  $m_K = 24$  (the range corresponding to a  $\delta = 10^{-4}$  transit of a  $m_K = 14$  star) is  $0.0086 \text{ stars-arcsec}^{-2}$ . The probability of any given small circle on the sky containing one of these stars is then simply the area of the circle multiplied by this density. Continuing this example, ( $m_K = 14, \delta = 10^{-4}$ ) the probability of such a star being within  $2''$  of a *Kepler* target star is 0.11.

However, because the *Kepler* field is quite extended and centered only a few degrees off the Galactic plane, there is a considerable gradient in background stellar density across the field that must be accounted for. To accomplish this, we simulate 21 different  $5 \text{ deg}^2$  fields, each centered on one of the *Kepler* double-CCD squares. The resulting probabilities are plotted in Figure 5.1 as a function of Galactic latitude, for the magnitude ranges corresponding to  $m_K = 11, 12, 13, 14$ , and  $15$ . Recognizing that this blend probability appears to be exponentially related to Galactic latitude  $b$ , and that the nature of the exponential depends on  $m_K$ , we fit an analytic expression of the following form:

$$p_{\text{blend}}(b, m_K) = C(m_K) + A(m_K)e^{-b/B(m_K)}, \quad (5.8)$$

where  $A$ ,  $B$ , and  $C$  are all polynomial functions of Kepler magnitude, with the coefficients listed in Table 5.1. These fits are valid between  $m_K$  values of 11 and 15, and  $b$  values between  $7^\circ$  and  $20^\circ$  (the approximate extent of the *Kepler* field). Figure 5.2 graphically illustrates the behavior of Equation 5.8.

### 5.2.1.2 Probability of an appropriate eclipsing binary

The probability that a blended star is an appropriately configured eclipsing binary system depends first on the binary fraction of blending stars, and secondly on both the distribution of binary properties and the magnitude of the *Kepler* target star. Of central importance is that in order for a blended binary to successfully mimic a *Kepler* planet transit candidate, it must have both a diluted primary eclipse shallow enough to look like a planet and a diluted secondary eclipse either shallow enough so as not to be detected or geometrically aligned so as not to occur.

The apparent fractional “transit” depth of a blended binary system depends on the intrinsic binary system eclipse depth  $\delta_b$ , and the relative apparent magnitudes of the *Kepler* target star and the blended system:

$$\delta = \delta_b \cdot 10^{-0.4(m_{K,\text{bin}} - m_{K,\text{target}})}. \quad (5.9)$$

The primary and secondary eclipse depths of the binary system are the following:

$$\delta_{b,\text{pri}} = \frac{\left(\frac{R_2}{R_1}\right)^2 F_1}{F_1 + F_2}, \quad (5.10)$$

and

$$\delta_{b,\text{sec}} = \frac{F_2}{F_1 + F_2}, \quad (5.11)$$

where  $R_1$  and  $F_1$  are the stellar radius and flux in the *Kepler* band of the larger of the two stars, and  $R_2$  and  $F_2$  are of the smaller star.

If the binary system has an eccentric orbit, then there is a chance that the orbit may be aligned such that only one eclipse occurs. The probability that *both* eclipses occur, given that at least *one* occurs, can be shown to be the following (using Equations 9 and 10 from Winn et al. (2010) and guided by Figure 3 in that work):

$$\begin{aligned} \text{Pr(both eclipses)} = \frac{1-e^2}{2\pi} & \left[ \int_0^\pi \frac{1-e\sin\omega}{1-e^2} d\omega + \right. \\ & \left. \int_\pi^{2\pi} \frac{1-e\sin\omega}{1-e^2} d\omega \cdot \frac{1}{\pi} \int_0^\pi \frac{1-e\sin\omega}{1+e\sin\omega} d\omega \right], \end{aligned} \quad (5.12)$$

where  $e$  is eccentricity and  $\omega$  is the argument of periastron (defined such that the transit occurs when the true anomaly  $f = \pi/2 - \omega$ ). The first term represents the probability that the one given eclipse occurs on the “apastron half” of the orbit (thus guaranteeing the other eclipse will occur); the second term is the probability that the given eclipse occurs on the periastron half, times the fraction of orbital orientations that give the second eclipse as well as the first. One can see that when  $e = 0$  this probability becomes 1. When  $e \rightarrow 1$ , the probability  $\rightarrow 4/\pi^2$ , or  $\sim 0.405$ .

The conditions we define for a binary to be “appropriate” are for the diluted primary eclipse depth to be between 0.02 and  $10^{-4}$  (shallow enough to look like a planet, but still detectable), and for the diluted

secondary to either be shallower than  $10^{-4}$  (undetectable) or not to occur. We recognize that “detectability” of a transit is a function of more than just the transit depth, but for our purposes we use a depth of  $10^{-4}$  as the detection threshold. A more detailed population study based on *Kepler* candidates should use rather the signal-to-noise ratio of a transit as the criterion for detectability (Beatty & Gaudi 2008a). However, as our framework deals with how to interpret signals once they are detected, careful detectability analysis is unnecessary.

To calculate the probability of all these conditions being met (a star being binary and being “appropriate”), we use the TRILEGAL simulations and assume binary properties according to the work of Raghavan et al. (2010). That is, we assume a flat mass ratio distribution between 0.1 and 1 (Raghavan et al. (2010) actually observes the distribution to be flat between about 0.2 and 1, but we extend it to 0.1 to be more conservative). We randomly assign eccentricities following the distribution in the Multiple Star Catalog (Tokovinin 1999).

For each star in a particular TRILEGAL line-of-sight simulation that lies in the appropriate magnitude range (§5.2.1.1), we first randomly assign it to be a binary or not and then calculate what the primary and secondary diluted depths would be if the system were eclipsing and blended with a *Kepler* target star of a particular magnitude.

$R_1$  and  $F_1$  are provided by TRILEGAL<sup>3</sup>, and we determine  $R_2$  and  $F_2$  based on a randomly assigned mass ratio and the Padova models at the age of the primary. Given these system parameters, we can then randomly determine if each system undergoes a non-grazing eclipse, according to the probability that each system will be in such an orientation:

$$\text{Pr}(\text{eclipse}) = \frac{R_1 - R_2}{a} \cdot \frac{1}{1 - e^2}, \quad (5.13)$$

where  $a$  is the orbital semi-major axis, determined from Kepler’s law, and  $e$  is the orbital eccentricity. We then determine whether one or both eclipses occur (according to Equation 5.12) and if only one occurs, then we randomly assign whether the lone eclipse is the “primary” or “secondary.”

From this procedure, using a *Kepler* target star of  $m_K = 14$ , an orbital period of 10 days, and a line-of-sight simulation at the center of the *Kepler* field, we find that 1.4% of binaries have non-grazing eclipses and about 27% of those eclipsing binaries are “appropriate.” For the binary fraction, we assume that  $\sim 40\%$  of stars have binary companions<sup>4</sup>, and then consider as potential false positives only the fraction of those binaries that are “short-period”; for our purposes,  $P < 300$  days, which comes out to about 1/8 of binary systems (again, according to the observed distribution of binary periods from Raghavan et al. (2010)). This gives an effective binary fraction of about 5%. This results in a probability of  $2.5 \times 10^{-4}$  for a star to be an appropriate eclipsing binary, giving a value of  $\pi_{\text{BB}} = 0.11 \times 2.5 \times 10^{-4} = 2.6 \times 10^{-5}$  for the center of the *Kepler* field.

As in §5.2.1.1, we empirically investigate how this probability changes as a function of galactic latitude and target star magnitude. We find the behavior for any particular magnitude is well described by a shallow

<sup>3</sup>This properly accounts for the possibility that the blend might be an evolved system; e.g. a dwarf star eclipsing a giant.

<sup>4</sup>To be precise, we actually use a binary fraction function that increases with stellar mass: 40% for  $M < M_\odot$ , 50% for  $M_\odot < M < 1.5M_\odot$ , and 75% for  $M > 1.5M_\odot$ , roughly adapted from Figure 12 in Raghavan et al. (2010). This is a conservative estimate of the binary fraction, as the Raghavan Figure includes multiple systems as well as binaries.

Table 5.1. Polynomial coefficients<sup>1</sup>for Equation 5.15

	$c_0$	$c_1$	$c_2$	$c_3$	$c_4$
$A$	-60.931	19.980	-2.4273	0.12912	-2.5038e-3
$B$	82.605	-25.320	3.0365	-0.15902	3.0668e-3
$C$	-0.27166	9.6434e-2	-1.2836e-2	7.5396e-4	-1.5465e-5
$D$	2.3345e-3	-7.1853e-4	8.1050e-5	-3.9902e-6	7.2607e-8
$E$	-4.7598e-2	1.5099e-2	-1.7333e-3	8.6394e-5	-1.5853e-6

<sup>1</sup>This table lists the polynomial coefficients for the empirical fits to how the blended binary false positive probability as a function of Galactic latitude changes with Kepler magnitude  $m_K$ .  $A, B, C, D$ , and  $E$  are functions of  $m_K$ , valid between  $m_K = 11$  and  $m_K = 16$ . The polynomials are of the form  $c_0 + c_1 m_K + c_2 m_K^2 + c_3 m_K^3 + c_4 m_K^4$ .

linear relation in  $b$ :

$$\text{Pr}(\text{appropriate ecl. binary}) = bD(m_K) + E(m_K), \quad (5.14)$$

where again the variation of the values of the coefficients  $D$  and  $E$  is modeled well with a polynomial in  $m_K$  (Table 5.1).

Multiplying Equation 5.14 with Equation 5.8 then gives a full analytic expression for the probability of a star of given Kepler magnitude at a given Galactic latitude to be blended with an eclipsing binary system able to mimic a planetary transit:

$$\pi_{\text{BB}}(m_K, b) = [C(m_K) + A(m_K)e^{-b/B(m_K)}] \times [bD(m_K) + E(m_K)], \quad (5.15)$$

where  $A, B, C, D$ , and  $E$  are polynomial functions of  $m_K$  with coefficients given in Table 5.1.

## 5.2.2 Hierarchical Triples

The probability that a *Kepler* target star is in fact a hierarchical triple system configured such that it might be able to mimic a planetary transit ( $\pi_{\text{HT}}$ ) can be broken down as follows:

$$\pi_{\text{HT}} = \text{Pr}(\text{triple}) \cdot \text{Pr}(\text{eclipsing and appropriate}). \quad (5.16)$$

The first factor is simply the frequency of triple systems. Raghavan et al. (2010) determine the frequency of multiple systems to be 12% for sun-like stars, and the Multiple Star Catalog (Tokovinin 1999) suggests about half of multiple systems have  $P < 300$  days, and so we adopt 6% as our triple fraction. The fraction of triple systems that are of appropriate configuration can be determined by using the same conditions as we used above in §5.2.1.2. That is, we require the diluted eclipse depths (Eqs. 5.9-5.11) to be between 0.02 and  $10^{-4}$  (again accounting for the possibility that an eccentric orbit might provide only a single eclipse), except



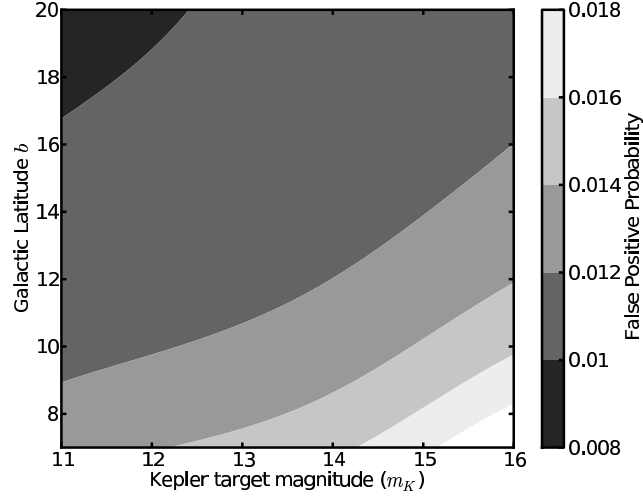


Figure 5.3 The false positive probability of a *Kepler* candidate, according to our basic framework (i.e. independent of  $\delta$ ), as a function of target star magnitude  $m_K$  and galactic latitude. A planet occurrence rate of 20% is assumed. This plot assumes that *Kepler* is able to internally restrict the radius inside which a possible blended binary might reside to  $2''$ . There is a small gradient across the field, but the false positive probability is uniformly low.

this time one of the three triple components provides the diluting flux.

We assume two different hierarchical possibilities for triple systems. Referring to the three components in order of descending mass as A, B, and C, the triple system may either be set up as A + BC, where B & C are the closer potentially eclipsing pair and A is the diluting star, or as AC + B, with A & C as the closer pair and B diluting. We ignore the case AB + C because the faintest component being the diluting star would be unable to mimic a planet transit.

We calculate the probability that a triple system will be eclipsing and “appropriate” (again assuming a 10-day orbit) as follows:

$$p_a = \int \int A(M_A, q_1, q_2) \Phi_q dq_1 \Phi_q dq_2. \quad (5.17)$$

$A(M_A, q_1, q_2)$  equals 1 if the system is eclipsing and can mimic a transit and 0 if not, and the mass ratios  $q_1 \equiv M_B/M_A$  and  $q_2$  (either  $M_C/M_A$  or  $M_C/M_B$ , with 50/50 odds) determine the architecture of the triple system.  $\Phi_q$  is the mass ratio distribution that we used in §5.2.1.2 (flat between 0.1 and 1). We assign the radius and flux of each component according to the Padova model grids in order to calculate both the non-grazing eclipse probability and the diluted eclipse depths, taking into account the effect of eccentric orbits. Evaluating this integral numerically we obtain  $p_a = 0.12$ , which results in  $\pi_{\text{HT}} = 9.8 \times 10^{-5}$ .

Unlike the blended eclipsing binary scenario, the probability of a target being a hierarchical eclipsing triple does not depend either on galactic latitude or apparent magnitude. There is a very weak dependence on stellar mass of the primary, but for our calculations we just assume that all target stars have masses close to  $1 M_\odot$ , which is reasonable as *Kepler* is specifically targeting solar-type stars.

### 5.2.3 Basic Framework: Summary and Discussion

Now that we have determined the priors for both false positive scenarios, we are able to evaluate the FPP (Equations 5.4 and 6.1) by assuming a frequency of close-in planets. We adopt a 20% frequency according to the results of the *NASA-UC Eta-Earth Survey* of Howard et al. (2010). This conservative estimate of 20%, combined with a 5% transit probability for a planet on a 10-day orbit (the period we have been assuming up to now) gives  $\pi_{\text{pl}} = 0.01$ , and thus an FPP of  $\sim 0.01$ . From a planet detection standpoint, this result is quite promising, as it gives a 99% probability that an observed planet-like transit signal around a  $m_K = 14$  star in the middle of the *Kepler* field is authentic. Because of the variation of the background stellar density across the field, this value varies with Galactic latitude and  $m_K$ , as shown in Figure 5.3. This is a remarkable result, as it indicates that almost every signal that passes the *Kepler* astrometric and photometric false positive tests is likely a planet transit, before any RV confirmation attempts.

One might rightly pause at this juncture and wonder how the false positive probability for *Kepler* can be so low. After all, transit searches up until now, both ground-based (e.g. HAT, WASP) and space-based (e.g. CoRoT), have been plagued by false positives (Konacki et al. 2003; O'Donovan et al. 2006; Poleski et al. 2010; Almenara et al. 2009). To address this, we consider what Equation 5.4 would say about the probability of a transit signal being true for those experiments.

Taking the Hungarian-made Automated Telescope Network (HATNet) as an example of a ground-based survey, we note that its 11cm telescopes produce a point-spread function of about  $14''$  in radius (Hartman et al. 2004), and thus a photometric aperture of  $\sim 25''$  in radius. Using this radius and a depth of 0.5% as a detection threshold, we repeat the analysis of §5.2.1, using the line-of-sight simulation at the center of the *Kepler* field for the sake of comparison. For the probability of a possibly blending star to be within the aperture, assuming a target star of 12th magnitude, we obtain 1.10, which must obviously now be interpreted as an average number of blending stars per aperture instead of a probability. For the probability of a blending star to be an appropriate eclipsing binary<sup>5</sup> we obtain  $9.7 \times 10^{-4}$ , giving  $\pi_{\text{BB}} = 1.10 \times 9.7 \times 10^{-4} = 0.0011$ . Following §5.2.2 we calculate  $\pi_{\text{HT}} = 2.1 \times 10^{-4}$ . Finally, taking into account that the probability of a sun-like star hosting a planet easily detectable by this survey is only about 1%<sup>6</sup>, then  $\pi_{\text{pl}} = 0.01 \times 0.05 = 5 \times 10^{-4}$  for this survey. This results in an FPP of 0.71 for a hot Jupiter-like transit signal for a HAT-like ground-based search, according well with Latham et al. (2009), who describe the results of follow-up efforts of a sample of transit candidates, eight of which turned out to be blended binaries and one to be a planet.

The space-based mission CoRoT (Baglin 2003) has also had difficulties with false positives. Though it obtains much better photometric precision than a ground-based search and benefits from uninterrupted observing, its large,  $320 \text{ arcsec}^2$  aperture (Almenara et al. 2009) results in an expected number of 2.98 blended stars for a  $m_K = 14$  target star (according to a simulation of one of the CoRoT lines of sight). In addition, its photometric precision is about one part in  $10^3$ , resulting in  $\pi_{\text{BB}} = 0.0031$ , and  $\pi_{\text{HT}} = 2.9 \times 10^{-4}$ . Assuming

<sup>5</sup>For both HAT and CoRoT, we relax the assumption that V-shaped transit signals are pre-vetted, so we allow for grazing blended binary orbits.

<sup>6</sup>for  $P < 11.5\text{d}$  and  $M > 0.5M_J$ ; Cumming et al. (2008).

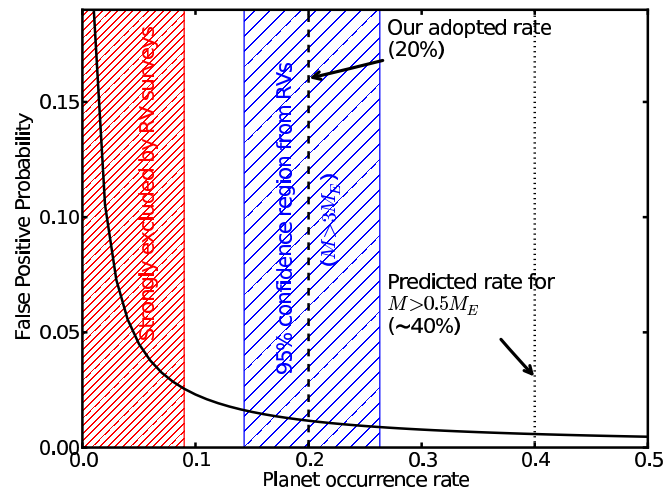


Figure 5.4 False positive probability as a function of assumed planet occurrence rate, for a  $m_K = 14$  target star in the center of the *Kepler* field. The occurrence rate of planets detectable by *Kepler* is not known for sure, but RV surveys, especially the *NASA-UC Eta-Earth Survey* of Howard et al. (2010), have made inroads in measuring the fraction of stars hosting low-mass planets. The hashed area below 9% represents the occurrence rate of planets with  $P < 50$  days that is ruled out with 95% confidence by  $\eta_{\text{earth}}$ , counting only the firm detections, and not correcting for completeness. The central hashed area represents the 95% confidence region calculated, including candidate planets and completeness correction, for minimum masses greater than  $3 M_{\oplus}$ . Extrapolating their observed mass distribution down to  $0.5 M_{\oplus}$  brings their total estimated planet occurrence rate to 43%. Overall, this plot shows that our derived FPP cannot reasonably be any higher than 3% if our planet occurrence estimate is incorrect, and will likely be lower.

then a 20% occurrence rate of planets detectable by CoRoT, this gives an FPP of 0.25. At first this appears to somewhat contradict Almenara et al. (2009), who reported 6 planets and 25 diluted binaries among CoRoT’s “solved candidates” (ignoring the “undiluted binary” category, as we are not considering that possibility for *Kepler*). However, if one considers how much easier (and faster) it is to identify a false positive than to positively confirm a planet, this prediction can certainly be consistent with these results, as only 49 of their 122 candidates had been solved at the time. In fact, a prediction of our methods is that many of the unsolved CoRoT candidates are indeed planets.

Another reasonable question to ask is how uncertainties in our models and assumptions propagate through to uncertainties in FPP. This is challenging to address exactly, as our analysis rests on the results from TRI-LEGAL simulations, stellar model grids, and various assumptions about multiple star systems. Rather than attempt a detailed start-to-finish treatment of all the uncertainties, we instead investigate what happens if we artificially inject fractional uncertainties into our prior calculations and simulate the results according to our analytic fits. We find that 20% fractional uncertainties in background stellar density, appropriate eclipse probability, and hierarchical eclipsing triple probability lead to 17% fractional uncertainty in FPP. This is a fiducial example, and the uncertainty in FPP scales linearly with these component uncertainties.

One might also wonder how sensitive our derived FPP for *Kepler* is to the assumption that 20% of stars

host planets, as well as how justifiable such an assumption may be. We address these questions in Figure 5.4. A 20% occurrence rate lies in the middle of the measured occurrence rate of planets with minimum masses  $> 3M_{\oplus}$  and periods  $< 50$  days from the *NASA-UC Eta-Earth Survey* of Howard et al. (2010). In addition, even the most pessimistic interpretation of the results from  $\eta_{\text{earth}}$  allows for a minimum of a 9% occurrence rate, which would still imply an FPP of only 3%. More likely, the true occurrence rate is somewhat higher than our assumption, if not as high as the  $\sim 40\%$  implied by a naïve extrapolation of the observed power law-like distribution down to  $0.5 M_{\oplus}$ . We note that the *NASA-UC Eta-Earth Survey*, as with all RV surveys, is only able to measure minimum masses and thus that the interpretation of the true mass of any individual detection is dependent on an assumption of the overall form of the planet mass function (Ho & Turner 2010b). However, when an ensemble of minimum mass measurements is available and its distribution resembles a power law with index  $\alpha < -1$ , the most likely explanation is that the true mass function follows a similar power-law shape.

In summary we may say that several factors contribute to *Kepler* being able to minimize the false positive problem compared to previous transit surveys. First, its ability to astrometrically rule out wide blend scenarios helps mitigate the issue of blended binaries. Secondly, its photometric precision enables it to identify many false positives based on their secondary eclipses. And lastly, *Kepler* is sensitive to lower-mass planets, which are significantly more common than the larger planets to which ground-based surveys are sensitive.

### 5.3 Detailed Framework: Considering Transit Depth

We note that we have not yet discussed any details of the transit signal besides its existence, though some of these details may be important. For example, one might expect false positive blended binaries to be more common at shallower depths (since faint stars are more common than bright stars, and thus more likely to be blended), which might make the BB scenario more of a problem for earth-sized transit signals. We have also assumed that planets and eclipsing binaries have the same eclipse probability (allowing us to cancel the likelihood factors in Equation 6.2), though this is not exactly true either, as both the orbital separations of the systems and the radii of the objects are different for a given fixed period. The eccentricity distribution of binaries is a function of period—longer period systems have larger eccentricities and thus false positive scenarios might be more likely to show only a single eclipse. And finally, for fainter stars and shallower eclipses, it may be more difficult for internal *Kepler* procedures to astrometrically identify blends.

With these concerns in mind, we may pursue a more detailed analysis of any particular transit. There are many features of transit light curves that might all be used in this exercise, but for now we only take into account the depth of the signal, as that is the most easily measured and easily understood quantity. In this case, Equation 6.2 becomes:

$$\text{Pr(pl}|\delta) = \frac{\mathcal{L}_{\text{pl}}(\delta)\pi_{\text{pl}}}{\mathcal{L}_{\text{pl}}(\delta)\pi_{\text{pl}} + \mathcal{L}_{\text{BB}}(\delta)\pi_{\text{BB}} + \mathcal{L}_{\text{HT}}(\delta)\pi_{\text{HT}}}. \quad (5.18)$$

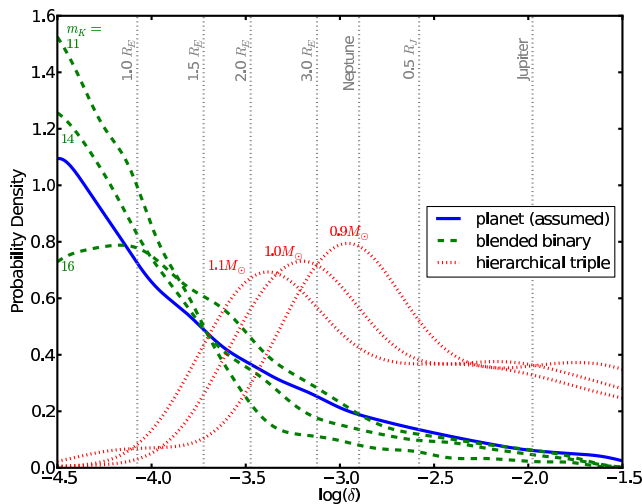


Figure 5.5 Distributions of apparent “transit” depths  $\delta$  for different scenarios. The blended binary and hierarchical triple distributions are based on TRILEGAL simulations with the binary distribution assumptions discussed in §5.2. Examples of  $\delta$  distributions are given for different target star properties, showing how the blended binary scenario depends on target star apparent magnitude and how the hierarchical triple distribution depends on intrinsic target star mass. The planet distribution comes from an assumption of a continuous power law in planet radius  $dN/dR_p \propto R_p^{-2}$ , including random statistical dilution by binary companions. Note how blended binaries become less significant for deep signals and how eclipsing triples become insignificant for the shallowest signals. The deep end of the false positive  $\delta$  distributions are mostly due to eccentric binaries that are oriented such that only a single eclipse occurs.

Here the likelihood functions provide a means to quantify the extent to which the conclusions of our simple framework may change as a function of transit depth  $\delta$ .

Figure 5.5 shows the likelihoods that we estimate for the three different scenarios as a function of depth. The distribution of depths for the blended binary and hierarchical triple scenarios are determined from the same calculations that we used to compute the priors, except rather than just counting all the systems that give depths that are both planetary and detectable, we keep track of the depth of each simulated false positive and build up  $\delta$  distributions.

We calculate the  $\delta$  distribution for planets assuming a simple continuous power law distribution of planet radii ( $dN/dR_p \propto R_p^{-2}$ ) between  $0.5$  and  $20 R_\oplus$ , and setting  $\delta = (R_p/R_\star)^2$ . While a more sophisticated treatment might involve adopting a planet mass distribution according to RV surveys and theoretical mass-radius relations (e.g. Fortney et al. (2007); Seager et al. (2007)), the number of assumptions required for these models and the fact that they do not generally include significant atmospheres for super-Earth-type planets suggests that such efforts are not warranted; for example, Rogers et al. (2011) show that Neptune-sized planets can have a wide range of possible masses. In addition, the current uncertainties in stellar radius of the *Kepler* candidate host stars further blur the mapping from  $\delta$  to  $R_p$ . Thus the main role of the  $\delta$  distribution we adopt for planets is to encapsulate the assumption that smaller planets are more common than large ones, which is

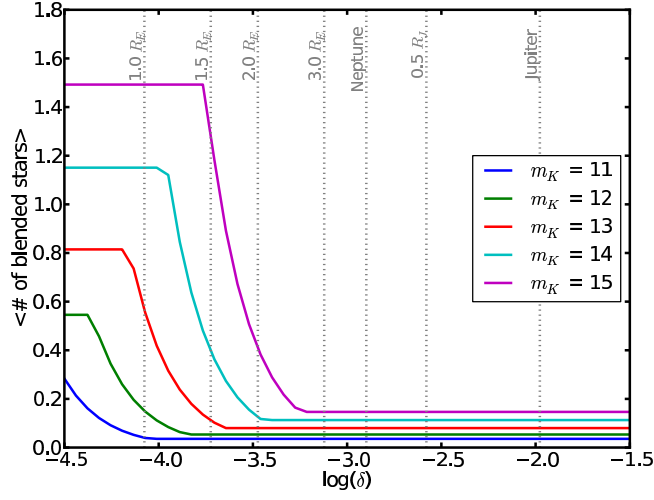


Figure 5.6 As stars get fainter and transit signals get shallower, the ability for *Kepler* to observe a centroid shift indicative of a displaced blended eclipsing binary decreases. We parametrize this effect according to Equation 5.19. The plateau towards shallow depths is a result of the maximum blending area for this example being set to an aperture of 8 *Kepler* pixels; the location of this plateau for any particular target will depend on its aperture size. This plot is made according to a Galactic latitude in the middle of the *Kepler* field; other latitudes will scale appropriately according to the varying stellar density. The planet radii are marked assuming a Solar-radius star.

consistent with radial velocity surveys (Howard et al. 2010).

Another consideration that should vary with  $\delta$  is the ability of *Kepler* to astrometrically identify displaced blends. In §5.2 we assumed a radius of  $2''$  inside which a blend might reside. However, this radius should increase as transits get shallower and stars get fainter, which should cause the signal-to-noise of the centroid shift signal to decrease. This is a question that the *Kepler* team should be able to address using simulations of its offset-detecting procedures, but for our purposes we use the radius that the *Kepler* team obtained for Kepler 10-b ( $1''.17$ ) (Batalha et al. 2011) and assume scaling with  $\delta$  and  $m_K$  as follows:

$$r = 1''.17 \sqrt{10^{-0.4(11-m_K)}} \left( \frac{\delta}{1.5 \times 10^{-4}} \right)^{-1}, \quad (5.19)$$

with 11 being the  $m_K$  value for Kepler 10. To be conservative we set the minimum  $r$  to be  $2''$  if this expression gives a smaller value. On the high end, we cap the radius at  $6''.4$ , corresponding an area equivalent to 8 *Kepler* pixels, a typical aperture size (though for any particular target this will vary). The square root factor accounts for a diminishing number of photons received as the target star gets fainter, and the inverse relationship with delta is because the centroid shift scales as  $\delta$ :  $\Delta C \sim \delta \cdot r$ . Figure 5.6 illustrates this effect; bright stars and deeper transits give Pr(blend) as determined in §5.2.1.1, but as the target star gets fainter and the signal shallower, the expected number of possibly blending stars begins to increase substantially, up to the point at which our calculated blend radius exceeds the maximum assumed 8-pixel aperture area.

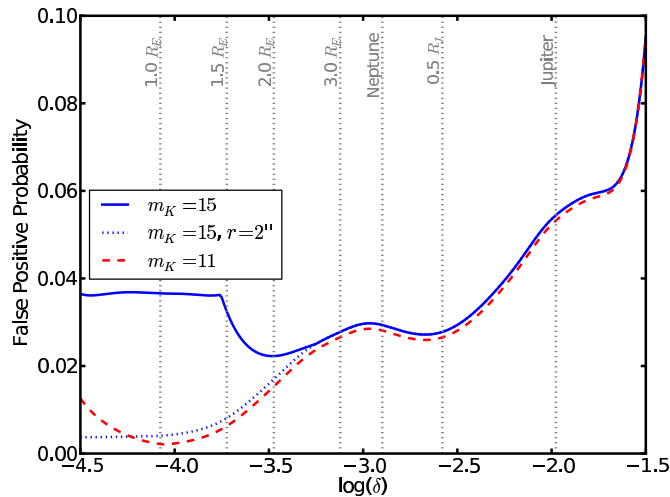


Figure 5.7 The probability that a 10-day signal of a given depth will be a false positive, shown for both an  $m_K = 11$  and an  $m_K = 15$  star with Solar properties, Galactic latitude in the center of the *Kepler* field, and an 8 pixel *Kepler* aperture. A overall planet occurrence rate of 20% and a planet radius function  $dN/dR \propto R^{-2}$  are assumed. The false positive probability increases towards deeper signals mostly because of the decreasing planet radius function combined with the significant tails of the false positive  $\delta$  distributions. For the fainter star the false positive probability begins to pick up again at the shallowest depths as it becomes more difficult for *Kepler* to rule out displaced blended binaries via astrometry. The plateau corresponds to the point at which the blending radius becomes equal to the aperture size. The dotted line represents the effect of restricting the blending radius to  $2''$  with a single high-resolution image; this can decrease the false positive probability for Earth-sized signals from  $\sim 4\%$  to  $< 1\%$ .

## 5.4 Results

### 5.4.1 General

The adoption of these more detailed considerations enables us to estimate the FPP as a function of  $\delta$  for a star of given apparent *Kepler* magnitude, Galactic latitude, stellar radius, stellar mass, and aperture size. This is illustrated in Figure 5.7 for fiducial examples of a signal with a 10-day period around  $m_K = 15$  and  $m_K = 11$  Sun-like stars in the middle of the *Kepler* field, both assumed to have 8-pixel apertures. We first note that over the whole range of  $\delta$  for these examples, the FPP remains  $< 0.10$ , indicating that these additional considerations do not significantly change the qualitative conclusions we reached within the simple framework. The majority of transit signals in the *Kepler* data release will be actual planets.

We note that FPP generally tends to *increase* with increasing signal depth. This may be understood by considering that the planet radius function decreases toward larger signals while the hierarchical triple false positive  $\delta$  distribution remains relatively flat (see Figure 5.5). Towards shallow signals FPP begins to increase again; this is due to the growing influence of the blended binary scenario as the radius inside which *Kepler* is able to rule out displaced blends decreases. The plateau at the shallowest depths is a result of the “blending radius” becoming equal to the size of the aperture, the maximum value it can attain. If this blending radius is

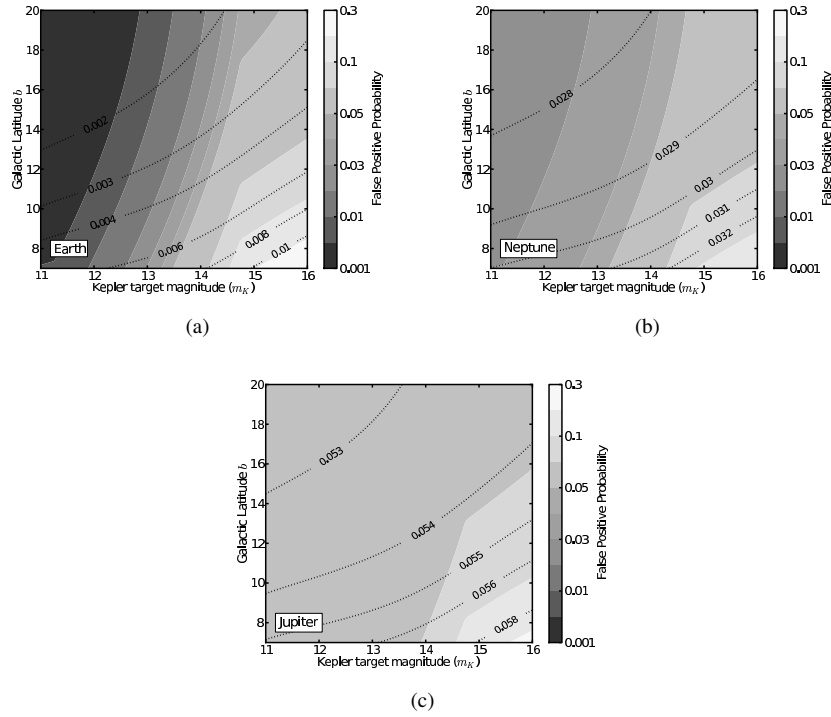


Figure 5.8 These plots illustrate the behavior of *Kepler* false positive probability (FPP) as a function of target star magnitude ( $m_K$ ) and Galactic latitude for three particular choices of transit depth  $\delta$ , all plotted with the same color scale. A planet occurrence rate of 20% is assumed, and the target star is fixed to have Solar mass, Solar radius, and a photometric aperture of 8 pixels. These plots are similar to Figure 5.3, except for that they take into account both the changing blend radius as a function of  $m_K$  and  $\delta$  (Equation 5.19) and the relative likelihoods of false positives and planets at the chosen values of  $\delta$ . All three  $\delta$  values show increasing FPP towards fainter target stars and lower galactic latitudes, though the strength of the gradient decreases for the deeper signals, as the relative importance of the hierarchical triple scenario increases. Dotted lines show the FPP contours if the blend radius were restricted to  $2''$ , illustrating the power of a single deep high-resolution observation. Note that only in the Earth-sized transit case does the high-resolution observation result in FPP  $< 0.01$ , as the FPP for shallower signals is dominated by chance-alignment blended binaries while the FPP for deeper signals is dominated by hierarchical triple scenarios.

able to be decreased to  $2''$  by a deep high-resolution image, that reduces the FPP at shallow depths to below 1%.

We note that the plots in Figures 5.5, 5.6, and 5.7 are only for particular chosen values of magnitude and a single Galactic latitude in the middle of the *Kepler* field, as well as for particular choices of stellar properties. We present a more comprehensive illustration of the FPP manifold in Figure 8, choosing three specific values of  $\delta$  to illustrate how the FPPs for different types of signals vary with target star magnitude and Galactic latitude. We fix the target star to have Solar properties in these examples.

Earth-sized transits show a steep gradient across the field and towards fainter stars; this is a result of increasing contribution to the FPP from blended binaries (see Figure 5.2), combined with the increased blend radius for a shallow transit (Figure 5.6). This gradient is shallower for a Neptune-sized signal and almost disappears for a Jupiter-sized signal, because of the growing contribution of the hierarchical triple scenario



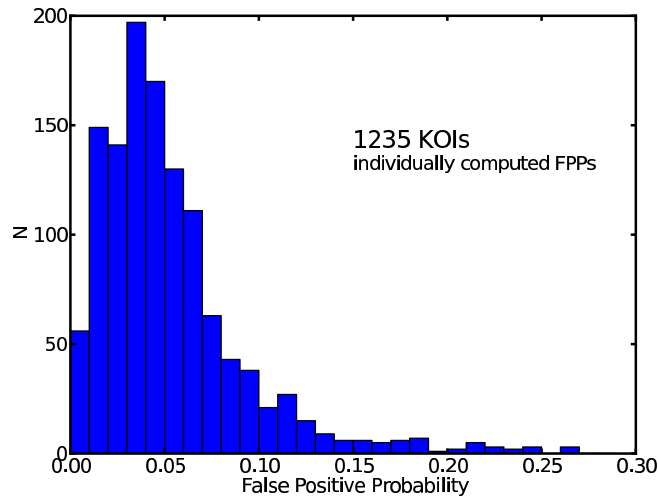


Figure 5.9 The distribution of false positive probabilities (FPPs) among the 1235 *Kepler* planet candidates announced in Borucki et al. (2011). FPP for each candidate is calculated individually, taking into account the apparent *Kepler* magnitude, Galactic latitude, mass and radius of the host star, the depth and period of the signal, the number of pixels contained the optimal aperture used for *Kepler* photometry, and the detection threshold of each KOI time series. Nearly 90% (1098) have FPPs less than 10%, and over half (713) have FPPs less than 5%. The mean FPP of the sample is about 6%, indicating that we expect there to be fewer than  $\sim 75$  false positives among the candidate sample. An important caveat here is that these calculations assume that all candidates have passed the false positive-vetting tests that are possible using *Kepler* photometry.

and decreasing influence of chance-alignment blended binaries. These plots also illustrate the potential power of deep high-resolution imaging follow-up observations. If such an image is taken and no companion is found outside a radius of a few arcseconds, then that dramatically reduces the FPP for shallow signals, as illustrated with the dotted contours.

## 5.4.2 Application to Kepler Candidates

We apply the framework discussed above to calculate the FPP for every *Kepler* Object of Interest (KOI) published in Borucki et al. (2011). For each KOI we generate individualized  $\delta$  distributions for the different false positive scenarios using the relevant *Kepler* magnitude, Galactic latitude, and stellar parameters from the *Kepler* Input Catalog. We also match the eccentricity distribution of false positive scenarios to the period of the signal (according to the multiple star period-eccentricity distribution as illustrated in Tokovinin (1999)) and use the actual detectability threshold for each KOI (based on calculating what depth corresponds to  $S/N = 3$  according to the data in Borucki et al. (2011)) to count “appropriate” eclipsing stellar systems, instead of the generic  $10^{-4}$  value. We then calculate the FPP using the actual area of the photometric aperture, as determined from the publicly available pixel data for each KOI and the transit depth as given in Borucki et al. (2011). The distribution of these calculated FPPs is illustrated in Figure 5.9.

## 5.5 Discussion: Relationship to “BLENDER”

The FPP analysis we present in this paper is not the first false positive analysis that has been done regarding *Kepler* candidates. In fact, the *Kepler* team has statistically “validated” three planets: Kepler-9d (Torres et al. 2011), Kepler-11g (Lissauer et al. 2011), and Kepler-10c (Fressin et al. 2011), by demonstrating that the chance of any of those signals being due to a false positive is low enough to consider the candidate a *bona fide* planet. This has been done using the procedure the *Kepler* team has named BLENDER (Torres et al. 2004, 2011).

BLENDER attempts to directly model the candidate light curve using every conceivable false positive scenario, informed by high-resolution imaging follow-up observations. The goodness-of-fit of the false positive models is then compared to the best-fit planetary model. The false positive scenarios that cannot fit the light curve as well as as a transiting planet model are rejected. The *a priori* likelihood of the remaining scenarios (those false positive scenarios that provide comparable-quality fits to the light curve) is then assessed relative to the likelihood of a *bona fide* transiting planet, and if the planetary explanation is much more likely, then the planet is considered validated.

As can be inferred from the fact that the *Kepler* team has published only three validated planets to date out of over 1200 planet candidates that have been made public, BLENDER is a very time-consuming procedure, being both computationally expensive and labor-intensive. Relying on extensive modeling of individual light curves and requiring a suite of follow-up observations to be most effective, it can only be applied to single KOIs on an individual basis.

If BLENDER may be characterized as a “deep and narrow” false positive analysis tool, the FPP analysis we present in this paper might be described as “shallow and wide.” It takes only 15 seconds<sup>7</sup> per candidate for us to generate the  $\delta$ -distributions required to calculate the individualized FPP numbers listed in Table ??, which makes our analysis easily and immediately applicable to all the KOIs, whereas BLENDER takes weeks of computation and analysis per candidate. On the other hand, BLENDER takes into account all possible information about each KOI (detailed light curve shape, AO imaging, multiwavelength transit information, etc.), whereas we only consider the depth and period of the transit signal and the properties of the target star.

Another way to think of the relationship between our FPP analysis and BLENDER is that if BLENDER is a  $N$ -step procedure, our analysis is step  $N$ . We ignore most of the detail of the light curve and make no use of any follow-up observations, but go straight to the *a priori* likelihood calculation and do that step as carefully as possible. What is remarkably encouraging for the *Kepler* mission is that even this “shallow” single-step analysis is enough to determine that the false positive probability for almost every KOI is less than 10%, and for over half the KOIs is less than 5%.

If our analysis is step  $N$  of the BLENDER process, how would the first  $N - 1$  steps be incorporated into the present analysis to improve upon the FPPs published here? First, consider that if  $x = L_{\text{pl}}$  and  $y = L_{\text{BB}} + L_{\text{HT}}$ ,

---

<sup>7</sup>non-parallelized computation on an iMac with a 2.66 GHz quad-core i5 processor

then the probability  $p_{\text{pl}}$  that a signal is a planet is the following:

$$p_{\text{pl}} = \frac{x}{x+y}. \quad (5.20)$$

This may be rewritten as

$$p_{\text{pl}} = \frac{1}{1+y/x}. \quad (5.21)$$

If  $y \ll x$  (as we have shown it typically is) then

$$p_{\text{pl}} \approx 1 - \frac{y}{x}, \quad (5.22)$$

or

$$\text{FPP} \approx \frac{y}{x}. \quad (5.23)$$

The typical role of BLENDER in this context can then be thought of as multiplying  $y$  by a factor we call  $f_{\text{BLENDER}}$  ( $0 < f_{\text{BLENDER}} < 1$ ) that represents the fraction of the potential false positive scenarios (weighted by their intrinsic likelihoods) that produce acceptable fits to the light curve. Thus if BLENDER were to rule out 90% of the false positive scenarios considered in our analysis ( $f_{\text{BLENDER}} = 0.1$ ) for a particular system, then that would decrease the FPP for that system by a factor of 10—such analysis would be enough to make  $\text{FPP} < 0.01$  for almost every KOI.

In a similar spirit, for those KOIs whose FPP is dominated by the blended binary scenarios (mostly the shallowest signals),  $y$  can also be significantly decreased simply if deep high-resolution imaging shows no potentially blending companions. This effect is illustrated in Figure 8, where dotted FPP contours are drawn illustrating the effect of restricting the “blend radius” to  $2''$ . Decreasing this even further to  $1''$  or smaller would give another factor of 4 or more reduction in FPP.

In some cases of course, follow-up imaging observations will identify the presence of nearby stars within the “blend radius” inside of which astrometric offset methods were previously unable to identify displaced blends. In these cases, the analysis presented in this paper must be superceded by a more specifically tailored analysis such as BLENDER. In general, a detected nearby blend will cause the preliminary FPP to substantially increase, as the  $\text{Pr}(\text{Blend})$  factor that we found to be of order  $\sim 0.10$  (§5.2.1.1) is then divided out from the  $L_{\text{BB}}$  term, making it more comparable to the  $L_{\text{pl}}$  term. In these cases a full suite of follow-up observations and the more detailed wholistic approach that BLENDER uses will become necessary to validate candidates.

## 5.6 Caveats and Conclusions

We present both a framework to analyze the *a priori* false positive probability (FPP) of *Kepler* planet candidates and preliminary FPPs for the entire sample of 1235 released candidates, finding that  $\text{FPP} < 10\%$  for almost all the KOIs and  $< 5\%$  for over half the KOIs. The philosophy we adopt in this work is to calculate

conservative upper limits to these FPPs; further analysis may well demonstrate them to be lower, but we do not expect them to be higher. Our analysis indicates that fewer than  $\sim 75$  are likely to turn out to be false positives.

However, the details of these conclusions are based on several assumptions (apart from those regarding the accuracy of the TRILEGAL simulations and the properties of binary and triple systems described in §5.2) that come with some caveats:

- We assume that all candidates have passed all preliminary false-positive-vetting procedures that are possible using *Kepler* photometry and astrometry alone. In particular we assume that the transits are not obviously V-shaped, that there is no detectable secondary eclipse, and that careful centroid analysis has not revealed the presence of a displaced blended binary. If photometry or astrometry for a candidate actually does turn out to indicate a possible false positive, then the FPPs calculated in this paper for that KOI are not accurate for that system.
- We assume host star stellar parameters according to the Kepler Input Catalog (KIC). If stellar radii or stellar types are found to be significantly different from the KIC estimates, then that could change the interpretation of transit signals (e.g. turning a Jupiter-sized planet into an M-dwarf).
- We assume a planet radius function that increases towards smaller planets. There are many reasons, both theoretical and observational, to assume this is correct, but if it is not, then the false positive numbers for the smallest candidates would be a factor of two or so higher.

We emphasize that in this analysis the only information about the transit that we have considered in detail is the depth. Certainly more information could be used, if one is willing to make more assumptions. For example, one might include a more detailed period dependence in the likelihood function by comparing the period distribution of known binary and triple systems to an assumed period distribution of planets, the way we have done with  $\delta$ . However, we do not believe that the true period distribution of exoplanets is known well enough yet for such an analysis to be useful.

We also emphasize that the intention of this paper is not to encourage other analyses to completely ignore the possibility that some *Kepler* candidates might be false positives. Rather, we suggest that in statistical analyses using the ensemble of KOIs to investigate the distribution of planet properties, the FPPs in this paper (or based on the calculations in this paper; e.g. with different assumptions of the planet occurrence rate or radius function) be used to count “fractional planets”; i.e. for a KOI with FPP = 0.05 to count as 95% of a planet.

Finally, we provide several suggestions to guide and optimize *Kepler* follow-up efforts, based on the results of our analysis:

- For the shallowest candidates, or those for which a blended binary is the most likely false positive scenario, we recommend deep high-resolution imaging (with a target contrast ratio corresponding to

the depth of the signal:  $\Delta m_K = -2.5 \log \delta$ ), as excluding the presence of potentially blending stars at close separation will be the quickest path toward validation of such systems. Contrast ratios up to 10 magnitudes as close as  $1''$  have long proven to be technically feasible with existing AO instruments (Luhman & Jayawardhana 2002; Biller et al. 2007, e.g.).

- For deeper candidate signals for which a hierarchical eclipsing triple is the most likely false positive scenario we recommend follow-up efforts targeted toward the identification of physically bound companions to the KOI. High-resolution imaging is one useful tool here (though not necessarily as deep as those observations targeting projected binaries) to target wide-separation physically-bound companions, but high S/N spectroscopy (both optical and infrared) may be even more important, in order to spectroscopically identify or constrain the presence of low-mass stellar companions.
- For all the candidates we recommend spectroscopic follow-up to improve our knowledge of the physical parameters of the candidate host stars, in order to rule out the possibility of an eclipsing binary being misclassified as a transiting planet due to an incorrect assumed radius.

In summary, the exquisite photometric and astrometric precision of the *Kepler* instrument enables many of the false positives that have traditionally plagued transit surveys to be identified prior to follow-up observations. The result is that the majority of the candidates announced by Borucki et al. (2011) are likely to be *bona fide* planets. Thus, having surveyed the landscape of false positives in the *Kepler* field, we conclude that the outlook is bright for statistical analyses of exoplanet occurrence and properties based on the data made public by the *Kepler* team.

## 5.7 Appendix: Blended Planets

In the present work we consider as false positives only astrophysical configurations that do not involve any planets but still mimic the signal of a transiting planet. However, there are various other scenarios involving “blended planets” that, while not strictly false positives (i.e. a transiting planet is still involved), may contribute significantly to uncertainty in the planet parameters derived from the transit signal. A “blended planet” for our purposes is a transit signal that appears to be a planet of a particular size transiting the target star but is actually a larger planet transiting a fainter blended star. As before, these scenarios can be divided into chance-alignment systems or physically associated hierarchical systems.

We have calculated that chance-alignment blended planets are significantly less likely to occur than their blended stellar binary cousins; this can be heuristically understood from the following considerations:

- Because the deepest intrinsic planetary transits have depths of only  $\sim 0.02$  and the diluted signal has to be detectable (we have adopted  $\delta \gtrsim 10^{-4}$  as a threshold), then the maximum contrast between the target star and the blending star is  $\Delta m_K = 5.75$ , which is significantly less than the  $\Delta m_K = 10$  we adopted

for blended binaries in §5.2.1.1. The sky density of stars available for the chance-alignment blended planet scenario is thus about 5.5 times lower than that for the blended binary scenario, according to the TRILEGAL simulations.

- Our assumed planet frequency ( $\sim 20\%$ ) is lower than our assumed binary fraction ( $\sim 40\%$ ).
- The largest planets, while the most amenable to causing the blended planet scenario because of their larger intrinsic transit depth, are the least common—only  $\sim 1\%$  of solar-type stars host close-in giant planets, and this occurrence rate is even lower for lower-mass stars (Endl et al. 2003; Johnson et al. 2010a), which are the most common blending stars.

Physically associated hierarchical planets, on the other hand, might well be relatively common compared to the stellar false positive scenarios or chance-alignment blended planets. Another way of saying this is that binary stellar systems are relatively common, and so it seems likely that a substantial fraction of *Kepler* targets (and therefore candidates) are in fact binaries of unknown architecture. The net effect of this on the interpretation of the sample of planet candidates will be additional uncertainty in the derived planet properties due to both diluting light from a binary companion and from possible stellar misclassification by the *Kepler* Input Catalog, which assumes each star is single. We note that the *Kepler* team does include blended planets in the BLENDER procedure, and in fact that such scenarios are often the most difficult to rule out (*Kepler* team, 2011, private comm.).

In summary, while the analysis presented in this paper may provide confidence that “classic false positive” stellar systems are not often masquerading as *Kepler* transiting planet candidates, we do caution that uncertainties regarding candidate host systems (including whether or not they are binary) must be considered in any statistical analysis of the whole candidate sample.

## Chapter 6

# An Efficient Automated Validation Procedure for Exoplanet Transit Candidates

*This chapter is adapted from “An Efficient Automated Validation Procedure for Exoplanet Transit Candidates,” ApJ 761,6, by TDM, and may be understood as a continuation and expansion of the ideas and methods presented in the previous chapter.*

### 6.1 Introduction

The first exoplanets found to transit the face of their host stars were all initially discovered by Doppler surveys, which detected minute radial velocity (RV) variations of stars indicative of the gravitational tug of orbiting planets (Charbonneau et al. 2000; Henry et al. 2000; Butler et al. 2004; Bouchy et al. 2005; Sato et al. 2005). Surveys designed to *detect* exoplanets via their transits soon followed (Alonso et al. 2004; McCullough et al. 2006; Bakos et al. 2007; Collier Cameron et al. 2007), motivated by the wealth of physical information that the light curve of a transiting planet can provide (Seager & Mallén-Ornelas 2003). While initial forecasts suggested that such programs would churn out planet detections at tremendous rates (Horne 2003), the actual yield of ground-based surveys has been much more modest, with only about a dozen discoveries published in the first five years of transit survey operations.

One of the major reasons for this slower-than-hoped-for discovery rate (in addition to the various contributing subtleties discussed in Beatty & Gaudi (2008b)) is the preponderance of various eclipsing stellar binary scenarios mimicking planet transit signals, or so-called “astrophysical false positives” (see §2.1), combined with the quantity and expense of the observational resources required to confirm a signal as truly planetary. The scale and scope of the *Kepler* mission multiplies this challenge to an unprecedented degree, as many of the stars and candidates are not easily amenable to traditional follow-up procedures.

Given these difficulties, it has become necessary to adopt a new paradigm of transiting planet confirma-

tion: *probabilistic validation*. The philosophy behind this strategy is to demonstrate that a particular transit candidate is much more likely to be a transiting planet than it is to be a false positive—which may be the only option available if positive confirmation via RVs or TTVs is impossible or impractical. This has been exemplified by the *Kepler* team’s BLENDER procedure, which is descended from the methods which helped identify OGLE-TR-33 as a false positive, and also used to help confirm other OGLE candidates (Torres et al. 2005, e.g.). Using a suite of follow-up spectroscopic and imaging observations combined with extensive light curve fitting, BLENDER rules out regions of parameter space where false positive models are unable to fit the photometric data as well as a planet model. First demonstrated with the validation of Kepler-9d (Torres et al. 2011), this procedure has been used to validate many of the planets which have so far received official *Kepler* number designations.

However, despite the successes of BLENDER, *Kepler* still has a problem of scale: BLENDER is expensive, both in CPU- and person-hours, and has not yet been applied to large numbers of candidates in a wholesale manner. As the overall mission of the *Kepler* project is not individual planet detections (as has been the case with ground-based surveys) but population statistics, there was a need to estimate the *a priori* false positive probabilities (FPPs) for *Kepler* candidates, before any follow-up observation or detailed BLENDER-style analysis. This was the inspiration for the work of Morton & Johnson (2011a), which demonstrated that for transit candidates that have passed the vetting tests that are possible with *Kepler* photometry alone, the FPP is typically around 5% or less.<sup>1</sup> For broad-brush studies to date, this result has been used to essentially ignore false positive contamination when analyzing the statistical properties of the candidate sample.

Despite the reassurance these *a priori* calculations provide that the broad characteristics of the overall *Kepler* candidate sample reflect those of the true exoplanet population, there remains significant motivation to push *a priori* FPP analysis beyond MJ11. First of all, MJ11 did not fully quantify all false positive scenarios; in particular, grazing eclipse signals, non-blended eclipsing binaries, and background blended transiting planets were not considered. A more thorough analysis will further increase the fidelity of the entire *Kepler* sample and thus enable more nuanced statistical analyses.

Secondly, the framework presented in MJ11 uses only minimal information about the photometric signals; that is, their depth and period. Much more information is available, particularly regarding the shape of the transit. In addition, MJ11 did not provide a transparent way to incorporate the results of follow-up observations to update the *a priori* calculations. And finally, as the process of probabilistically validating a planet is the same as demonstrating that its FPP is sufficiently low, a more stringent FPP calculation in the style of MJ11 may be used to validate large numbers of transit candidates for which traditional confirmation is difficult or impossible. In addition to enabling ground-based surveys to streamline follow-up efforts, such

---

<sup>1</sup>The dominant reason why this is much lower than has been the case for ground-based surveys is because the quality of the *Kepler* photometry allows for much more constraining pre-followup vetting than has been possible with ground-based surveys; in particular, *Kepler* is able to identify many blended binary scenarios by measuring the target’s center-of-light to shift during eclipse. This makes the effective “blend aperture” for *Kepler* much smaller than for ground-based wide-field surveys. Additionally, the precise photometry can put strict limits on the presence of any secondary eclipse, which further constrains false positive scenarios. And finally, the signals are typically shallower than detected in ground-based surveys, which also contributes to intrinsically lower FP rates.



a tool could ensure that the legacy of the *Kepler* mission is indeed thousands of transiting planets, not just planet candidates.

This paper presents a comprehensive transit candidate validation procedure that is based on the *a priori* framework of MJ11, yet also uses information about the shape of the transit signal and naturally incorporates (yet does not require) follow-up spectroscopic and imaging observations. An early version of this analysis, representing a halfway point between MJ11 and this work, was used to validate the KOI-961 planets (Muirhead et al. 2012b). Crucially, this method is capable of analyzing large numbers of candidates quickly; the entire end-to-end computation for a single transit candidate signal takes only about 10 minutes on a typical personal computer. It is thus capable of being applied to the entire *Kepler* candidate sample in the immediately foreseeable future. Looking beyond *Kepler*, the larger goal of this procedure is to revolutionize the follow-up strategy of large-scale transit surveys by reducing the time and cost necessary to identify false positives or confirm transiting planets, and to enable statistical analyses of transit candidate populations without having to wait for every candidate to be individually solved.

In §6.2 I review the probabilistic framework introduced in MJ11 and summarize the false positive scenarios considered in this work. I describe the entire procedure in detail in §6.3, present the results of applying this procedure to a number of previously studied *Kepler* signals in §6.4, discuss the relationship of this work to MJ11 in §6.5, and provide concluding remarks in §6.6.

## 6.2 Framework

Validating a transiting planet means demonstrating that the false positive probability (FPP) of the signal is small enough to be considered negligible. Regardless of exactly where this threshold lies, the process of validation is the same as carefully calculating the FPP. Following MJ11, I define the FPP for a given transit signal:

$$\text{FPP} = 1 - \Pr(\text{planet} \mid \text{signal}) \quad (6.1)$$

where the vertical line means ‘given’ (i.e. “the probability of there being a planet given the observed signal”), and

$$\Pr(\text{planet} \mid \text{signal}) = \frac{\mathcal{L}_{\text{TP}}\pi_{\text{TP}}}{\mathcal{L}_{\text{TP}}\pi_{\text{TP}} + \mathcal{L}_{\text{FP}}\pi_{\text{FP}}}. \quad (6.2)$$

In the above equation,  $\mathcal{L}$  represents the Bayesian likelihood factors (the “probability of the data given the model”), and  $\pi$  represents the priors (the *a priori* probabilities that each given scenario exists). The TP subscript represents the true transiting planet scenario, or “true positive,” and FP represents all of the false positive scenarios.

The false positive scenarios I consider in this analysis are the following:

- A non-associated foreground or background eclipsing binary system is blended within the photometric

aperture of the target star (BEB).

- The target is a hierarchical triple system in which two of the components eclipse (HEB).
- The target star is an eclipsing binary (EB).
- A non-associated blended foreground or background star happens to have a transiting planet (Bpl).

This list is more comprehensive than MJ11, which only considered BEBs and HEBs quantitatively, and comprises every conceivable astrophysical false positive scenario.

## 6.3 Procedure

The validation procedure I present in this paper has the following four steps:

1. Simulate a **representative population** for each scenario, fixing the period to be that of the transit candidate signal.
2. Use this population to calculate the prior for each scenario, optionally taking into account any follow-up observations that may exist.
3. Use the same simulated population to calculate the likelihood of the observed transit signal under each scenario.
4. Combine these numbers to calculate the FPP (Eq. 6.1) of the signal under an assumption of the **specific planet occurrence rate**; if this number is significantly  $< 1\%$  for a conservative estimate of planet occurrence, then consider the planet validated.

In the following subsections I explain each of these steps in detail, including the important terms in boldface above.

### 6.3.1 Population Simulations

Central to the validation procedure I present in this paper is the idea of simulating “representative populations.” This means using physically and observationally informed assumptions to generate a population of different instances of a particular transit or eclipse scenario at a given fixed period, such that the simulated population accurately reflects the true population. In other words, a representative population must be created such that each instance in the population is equally likely to exist; this makes averaging over the population trivial, as each instance has equal weight in defining the distribution of any desired quantity. For clarification, in this paper the following boldfaced words conform to this usage: a **representative population** for a particular **scenario** (e.g. BEB) is made up of many simulated **instances** of that scenario. For all the results presented in this paper, I create each representative population to have 20,000 instances.

These simulations necessarily require several assumptions:

- The population of stars in a cone around the line of sight.** This requires assumptions about both Galactic structure and stellar evolution. TRILEGAL (TRIdimensional model of the GALaxy; Girardi et al. 2005, and used in MJ11) is a tool perfectly suited to this purpose. TRILEGAL takes as input a given direction and area on the sky, a value for extinction at infinity, and parameterizations of the Galactic structure (e.g. scale height of the disk), and returns an appropriately simulated stellar population, with apparent magnitudes in any chosen photometric system as well as the physical properties of all the stars.
- The properties of multiple star systems.** This encompasses both multiplicity fractions and the distributions of physical and orbital properties for multiple systems, such as the mass ratio and eccentricity distributions. All these properties may be adopted based on the statistics of observational multiplicity surveys, such as Raghavan et al. (2010). To simulate the physical distributions of multiple systems, I first start with a total mass of a star system, split it into two components with 50% probability, and then split one of those into two more with 25% probability. Each time I make a split I choose a mass ratio uniformly between 0.1 and 1. There is no specific physical motivation for this procedure, but it reproduces well the observations of Raghavan et al. (2010), who observed roughly 35-40% of stars to be in binary systems and about 12% to be in triple or higher-order multiple systems, and a flat mass ratio distribution for binary pairs. Besides the binary fraction, I also must assume a “close binary fraction,” which is the fraction of stars that have companions with periods small enough to mimic the observed planet transit candidate population. For this cutoff I choose  $P < 300$  days; according to the period distribution observed by Raghavan et al. (2010) this encompasses about 12% of binary stars. Therefore I assume a “close binary fraction” of  $0.05 \approx 0.4 \times 0.12$ . Unlike the binary fraction, we do not use a pared-down “close hierarchical fraction,” under the assumption that most systems in triple and higher-order systems will undergo significant orbital evolution over their lifetimes, ending up with at least one close pair, and according to studies that have observed this (Tokovinin et al. 2006).
- Stellar models.** Accurately simulating eclipsing star systems requires assigning appropriate physical and observational properties to simulated binary companions. Given a primary star of a particular mass, age, and metallicity, a simulated binary companion is assigned a mass according to the assumed mass ratio distribution, and then a radius and magnitudes in different passbands according to stellar model predictions for that particular mass, age, and metallicity. Models are also required to accurately simulate a target star as a hierarchical system, which requires estimating absolute magnitudes for the target star in addition to its companions, in order to calculate both blended eclipse depths and the total colors of the system. In this work I use the Padova stellar models (Girardi et al. 2002) for all masses  $> 0.15 M_{\odot}$  (the lower limit returned by the online interface that provides the models) and the Dartmouth models (Dotter et al. 2008) below  $0.15 M_{\odot}$ . Preference is given to the Padova models for the sake of consistency, because they are used to populate the TRILEGAL simulations. Below  $0.11 M_{\odot}$  (the lower

limit for the Dartmouth models), I use the models of Baraffe et al. (2002).

- **Planet occurrence rate.** Somewhat paradoxically, in order to complete the FPP calculation for any given transit signal, prior assumptions about the planet occurrence rate and radius distribution must be made. See section 3.4 for how I treat this delicate issue by introducing the concept of the **specific occurrence rate**. For the background transiting planet scenario, I use a generic assumption of a 40% planet occurrence rate, with a power law distribution of planet radii  $dN/dR \propto R^{-2}$ .

Due to all these assumptions, the procedure I present in this paper is model-dependent. However, this is a feature of all Bayesian analysis, which always depends on assumptions (priors) and models, whether physical or empirical. The advantage of the Bayesian approach is that it provides a natural way to incorporate the existing body of knowledge into a new analysis, and to make assumptions explicit. That is, there has been significant work in astronomy dedicated to investigating Galactic structure, stellar models, and stellar multiplicity, and the strategy of this analysis is to attack this probabilistic validation problem armed with as much of this prior knowledge as possible.

### 6.3.2 Priors

I define the prior ( $\pi_i$ ) for each scenario to be the product of three distinct factors: the probability that the astrophysical scenario in question exists within the photometric aperture, the geometric probability that the orbit is aligned such that an eclipse is visible, and the probability that the eclipse is “appropriate”—that is, able to mimic (or be detected as) a transiting planet. I calculate the first of these factors based on the Galactic population and stellar multiplicity rate assumptions discussed in §6.3.1, combined with knowledge of the sky area inside which a FP might reside.

The probability of an eclipse occurring, given random orbital orientations with respect to the plane of the sky, is

$$\text{Pr}_{\text{ecl}} = \left( \frac{R_1 + R_2}{a} \right) \left( \frac{1 \pm e \sin \omega}{1 - e^2} \right), \quad (6.3)$$

where  $R_1$  and  $R_2$  are the radii of the two bodies in question;  $a$ ,  $e$ , and  $\omega$  are their orbital semimajor axis, eccentricity, and argument of periastron; and the  $\pm$  symbol is a  $+$  for primary eclipse and  $-$  for secondary.<sup>2</sup> Calculating this probability for a particular scenario requires averaging over the distribution of the various physical and orbital quantities of that particular scenario (i.e. larger stars are disproportionately likely to host a binary star that eclipses, for a given period); in practice this is accomplished simply by simulating a representative population (§6.3.1) for a particular scenario, assigning isotropic orbital inclinations, and counting what fraction result in an eclipse (either primary, secondary, or both).

<sup>2</sup>Note that using the sum of the radii in this equation for transit probability accounts for the possibility of grazing eclipses; this is different from MJ11, who used the difference of the radii, explicitly rejecting grazing eclipses as potential false positives, on the grounds that any explicitly V-shaped transit could be vetted as a probable false positive based on photometry alone. In this work, I include the possibility of grazing transits in all the scenarios.

In order for an eclipse of a certain scenario to be “appropriate,” it must pass all vetting constraints available, and the probability that a certain scenario will result in an appropriate eclipse may be calculated by just counting what fraction of a representative population pass the constraints. For the analysis in MJ11, the only constraint we used was depth: the diluted primary eclipse depth must be deep enough to be detected, while the secondary eclipse depth must be shallow enough *not* to be detected, given a detection threshold (or one of the eclipses must be missing due to an eccentric orbit). In the present work, I have generalized the framework in order to naturally include different types of supplementary observations—either preparatory or follow-up, as the case may be.

For example, as part of the simulations, I assign each false positive instance apparent magnitudes in many passbands. This allows observed colors of the target star to be used to constrain the population simulations. For example, in the case of *Kepler* targets, each star has *Kepler* Input Catalog (KIC) *griz* + 2MASS *JHK* magnitudes, and we constrain our representative populations of HEB and EB scenarios to conform to the observed total  $g-r$  and  $J-K$  colors, to within 0.1 mag (I choose these two colors in particular, and not  $g-K$ , for example, to avoid being too sensitive to metallicity effects (Johnson & Apps 2009)). This results in asymmetric distributions of target star properties, as shown in Figure 6.1. If a high-resolution spectrum is obtained for the target star, allowing for spectroscopic measurements of  $T_{\text{eff}}$  and  $\log g$  of the star that dominates the system luminosity, then these populations may be constrained to require the primary star to have the measured properties.

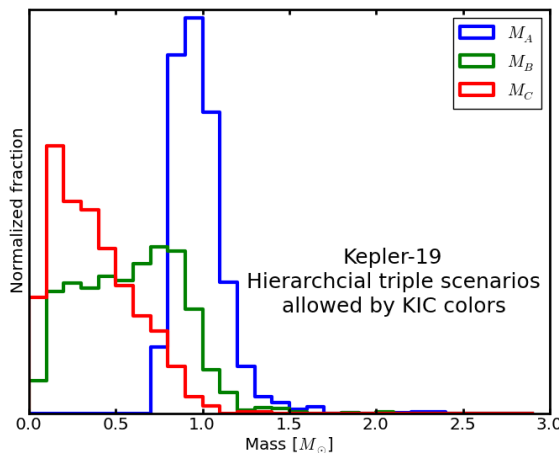


Figure 6.1 The distribution of hierarchical triple systems allowed for Kepler-19b (KOI 84.01) by the KIC  $g-r$  and 2MASS  $J-K$  colors. Note how there is a tail of higher-mass systems that can mimic Solar-like colors. This is important for calculating FPP, because systems with higher-mass (and thus higher-luminosity) primaries are more prone to cause false positives with shapes similar to planet transits. If the spectroscopic properties of the primary are measured (e.g. with a high-resolution spectrum), then this significantly constrains the false positive landscape (see Figure 6.6).

In addition, for each blended false positive instance I assign an angular separation from the target star on the sky, assigned differently for different scenarios. An important quantity for this step is the so-called

“confusion radius,” or the separation from the target star inside which a blend might reside. This may be constrained either by centroid analysis or simply by the size of the aperture, to be most conservative. Chance-aligned blends are assigned positions perfectly uniformly within the confusion radius surrounding the target star, whereas the positions of hierarchical scenarios are assigned according to the assumed orbital distributions and random mean anomalies.

Using these separations in combination with the multi-band photometry information, I may use a contrast curve (magnitude contrast as a function of radius from the target star) that results from analyzing a high-resolution follow-up image in any passband to rule out all instances of a particular scenario that *could have been detected* by the observation. The same principle may be naturally used to rule out instances based on other types of follow-up, such as multi-color transit observations testing for color-dependent transit depth. After including all available supplementary observational information—or none, as the case may be, in which case the only constraints are the primary/secondary depth conditions and the confusion radius—the “appropriate probability” is then the fraction of instances of a particular scenario that pass all the vetting tests.

### 6.3.3 Likelihoods

Once the representative population for a given scenario is simulated and vetted according to whatever information is available, then the likelihood  $\mathcal{L}$  for a given transit signal under that scenario may be calculated. It is at this stage in the procedure that the method introduced in this paper advances most significantly beyond the preliminary work of MJ11, so I will first give a brief review of how we treated likelihoods in that work.

In a Bayesian calculation, the likelihood factor may be colloquially described as “the probability of the data given the model.” In the simplest case, if the data consists of a single measurement, and the property being measured follows a known probability distribution, then the likelihood of the data is simply the known probability density function (PDF) of the property evaluated at the location of the measurement. This is exactly how we treated the likelihood factor in MJ11: the “datum” was the depth of the transit signal, and the PDF was determined using the distribution of depths from the representative population simulations.

The crucial step forward I take in this work is to take full advantage of fact that there is much more information in the shape of a transit signal than just its depth. In short, transiting planet and diluted eclipsing binary signals, at a fixed period, generally do not look the same. This is the starting point and backbone of the *Kepler* team’s BLENDER analysis, in which the goodnesses-of-fit of exhaustive simulations of putative false positive light curve models are compared to the goodness-of-fit of a transiting planet model; false positive models that do not produce acceptable fits are ruled out.

In contrast to BLENDER, I do not fit physical models to light curves to rule out false positive scenarios. Instead, I use the transit signal shape by extending the likelihood analysis of MJ11 into two additional dimensions: duration and slope of the signal, in addition to depth. I accomplish this by using our representative population simulations to define the three-dimensional probability distribution of these transit shape parameters, and then evaluating this PDF for the shape of the observed signal—the exact three-dimensional

analogue of the likelihood calculation of MJ11. To enable this, I turn to the simplest of transit shape models: the trapezoid.

For each instance in the representative population of each scenario, I follow a two-step procedure. First, I model the exact shape of the eclipse (accounting for every detail of the eclipsing system including the eccentricity and orbital orientation provided by the simulation) using a limb-darkened eclipse model (Mandel & Agol 2002) with appropriate quadratic coefficients (Claret 2000), taking into account the 30 minute *Kepler* long-cadence integration time. Secondly, I fit that physically modeled light curve with a simple trapezoid function. I parameterize the trapezoidal shape with three parameters: depth, total duration (from first to last contact), and the ratio of total duration to ingress time, or “slope,” which parametrizes the shape of the signal: i.e. whether it is “box-shaped,” “V-shaped,” or somewhere in between. From each population I thus obtain a three-dimensional “scatter plot” of shape parameters, which I turn into a three-dimensional PDF using Gaussian kernel density estimation (using the `gauss_kde` routine from the `scipy.stats` Python library<sup>3</sup>).

Once this PDF has been created for each scenario, the each likelihood may be calculated as well. To this end, I fit the trapezoidal shape to the observed phase-folded transit light curve, using a Markov Chain Monte Carlo routine (Foreman-Mackey et al. 2012) in order to obtain reliable error bars (important for signals with lower signal-to-noise ratios). From the MCMC chains I construct a three-dimensional posterior PDF, and then calculate the likelihood by multiplying the two PDFs (“model” and “data”) together and integrating over all three dimensions. This is just like the simple likelihood calculation discussed above, except that instead of evaluating the model PDF at a single point, I integrate over the uncertainty in the data values. The procedure is illustrated in the figures in §6.4, which also give a sense of what the 3D shape distributions look like for different scenarios.

Note that nowhere in this process do I attempt to derive any physical properties (e.g. impact parameter,  $a/R_*$  etc.) from the trapezoidal fit, nor do I at any point fit a physical transit model to the observed data—the trapezoid function is purely descriptive, and precise physical properties of the transiting planet are unnecessary in order to calculate its FPP. I also do not use the candidate’s period in the likelihood analysis, apart from the fact that all the simulated scenarios are simulated to have the same period as the candidate. In other words, I do not attempt to compare the period distributions of the various false positive scenarios with an assumption regarding the planet period distribution. The only question I aim to answer with this likelihood calculation is “How much does this signal look like each scenario?” and ultimately “Does this signal look much more like a transiting planet than a false positive?”

---

<sup>3</sup>[http://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.gaussian\\_kde.html](http://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.gaussian_kde.html)

### 6.3.4 Final Calculation

Besides the priors and likelihoods for all the false positive scenarios, calculating the FPP for a transit signal requires assumptions about the true distribution of transiting planets. Here my strategy departs from MJ11. In MJ11, we calculated FPPs under the assumption of an overall 20% planet occurrence rate, and a distribution of planet radii  $dN/dR \sim R^{-2}$ . In order to avoid such blanket assumptions of planet occurrence rates and radius distributions in this work, I proceed as follows.

Rather than use a generic distribution of planet radii in order to simulate the representative population of transiting planets that I need (in order to determine the TP prior and likelihood necessary to complete the FPP calculation), I instead simulate a distinct “radius bin” population, customized to the signal in question. In other words, I draw planet radii from a uniform distribution between  $2/3$  and  $4/3$  the best-fitting planet radius and simulate the representative population of transit signals using these radii. Of course, this nearly always makes the depth of the signal lie very near the middle of the depth distribution; this is by design. If the duration and slope parameters also lie near the middle of the TP distribution while being on the outskirts of all the FP distributions, this may by itself be evidence enough to validate the planet, as I demonstrate in §6.4.

While simulating the planet population to have just the right size to match the observed signal may seem to give an improper advantage to the planet model, this is perfectly reasonable as long as the prior is treated appropriately. No longer may a generic planet occurrence rate (e.g. “40% of stars have planets”) be used; instead, the relevant occurrence rate is the assumed fraction of stars that have planets *in the radius bin in question*. I call this the **specific occurrence rate**  $f_p$  and define it as an integral over the planet radius distribution function  $\Phi_R$ :

$$f_p(R_p) = \int_{0.7R_p}^{1.3R_p} \Phi_R(R) dR \quad (6.4)$$

What sort of assumption is reasonable for this  $f_p$ ? For example, if  $\Phi_R(R) \propto R^{-2}$  from 0.5 and 20 Earth radii, 34% of the probability lies between 0.7 and 1.3  $R_\oplus$ , whereas only 3.4% of the probability lies between 7 and 13  $R_\oplus$ . Thus if an overall 40% occurrence rate were assumed with this radius distribution, the specific occurrence rate of a  $\sim 1 R_\oplus$  planet should be  $\sim 14\%$  and the specific occurrence rate of a  $10 R_\oplus$  planet should be 1.4% (see Figure 6.2 for more illustration).

However, the motivation for this “radius bin” strategy is to avoid relying on an assumption of an overall planet occurrence rate and radius distribution, and integrating a generic power law to obtain a specific occurrence rate does not accomplish this. Instead, consider the following rearrangement and simplification of Equation 6.2:

$$\text{FPP} = \frac{L_{FP}}{L_{FP} + f_p L_{TP}}, \quad (6.5)$$

where

$$L_{FP} = \sum_i \pi_i \mathcal{L}_i, \quad (6.6)$$



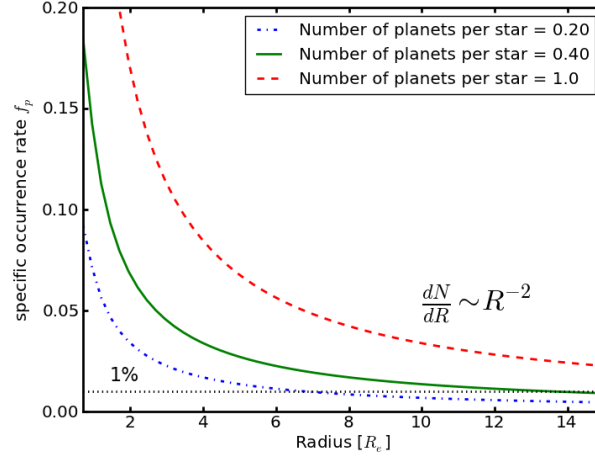


Figure 6.2 The specific occurrence rate  $f_p$  as a function of radius for three examples of generic radius distributions, with all scaling as the inverse square of the planet radius. I define the specific occurrence rate at a particular radius  $R_p$  to be the integral of the planet radius probability density function between  $0.7R_p$  and  $1.3R_p$  (Equation 6.4). Even though the use of the specific occurrence rate in the analysis presented in this paper makes it independent of assumptions of the overall planet rate or the exact shape of the radius distribution, this illustrates what reasonable values of  $f_p$  might be. For example, it is clear that  $f_p = 0.01$  is quite conservative for nearly any planet size, especially for small planets.

with  $i$  representing all of the false positive scenarios,  $f_p$  being the true specific occurrence rate, and  $L_{TP}$  equalling  $\pi_{TP}\mathcal{L}_{TP}$  if the specific occurrence were unity (i.e. assuming every star has a planet in this radius bin). Equation 6.5 may be further simplified as

$$\text{FPP} = \frac{1}{1 + f_p P}, \quad (6.7)$$

where  $P = L_{TP}/L_{FP}$ . This allows me to parametrize the FPP in terms of the unknown specific occurrence rate  $f_p$  and the factor  $P$ , which is independent of assumptions about either planet occurrence or the true planet radius distribution.

The final result of this validation method may thus be presented as simply this factor  $P$ , which contains all the information necessary to calculate the FPP based on an assumed specific occurrence rate (Equation 6.7). This lends ease to interpretation, as for large values of  $P$ , the FPP is very nearly  $1/(f_p P)$ . Thus if  $P = 1000$ , then under an assumption of  $f_p = 0.5$ , the FPP is 1 in 500. For  $f_p = 0.2$ , FPP would be 1 in 200, and so on.

Another useful presentation answers the following question: at what value of the assumed specific occurrence rate would the FPP be low enough to consider the planet validated? This is simply a matter of choosing a target FPP (e.g.  $\text{FPP}_V$  for “validation FPP”) and solving for  $f_p$  in Equation 6.7:

$$f_{p,V} = \frac{1 - \text{FPP}_V}{P \cdot \text{FPP}_V} \quad (6.8)$$

which may be approximated by  $f_{p,V} = (P \cdot \text{FPP}_V)^{-1}$ . For example, for a target FPP of 0.5%,  $P = 1000$  requires

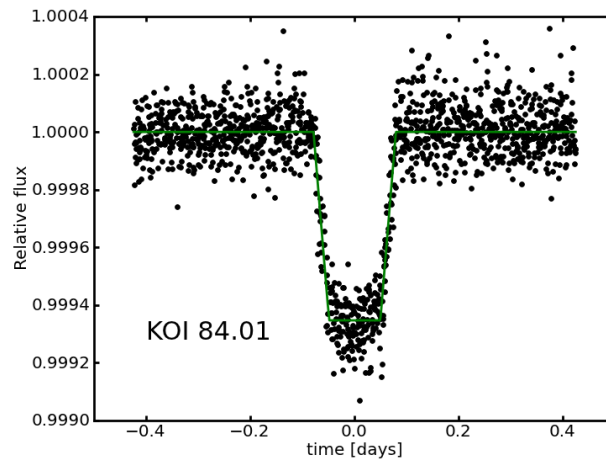


Figure 6.3 The phase-folded *Kepler* photometry of Kepler-19b (KOI 84.01). The solid line illustrates the best-fitting trapezoid model used for the likelihood calculations (§6.3.3). Note that the transit is clearly box-shaped, not V-shaped.

$f_{p,V} = 0.2$  to validate, while  $P = 2 \times 10^4$  requires only  $f_{p,V} = 0.01$ . Recent results from both RV surveys (Howard et al. 2010) and *Kepler* (Howard et al. 2012) suggest that over 40% of stars have planets of some sort within  $\sim 50$  day orbits, and that the frequency of planets rises with decreasing radius. Other work indicates both that the average number of planets per star may even be  $>1$  (Youdin 2011) and that fraction of stars with planets may also be close to unity (Cassan et al. 2012). Thus, if  $f_{p,V}$  for a particular transit signal is measured to be significantly below the curves plotted in Figure 6.2 (e.g. if it is 2-3% for a signal of a  $2R_{\oplus}$  transit), the planet may be considered securely validated.

## 6.4 Tests

In this section I present the results of applying this procedure to a sample of *Kepler* Objects of Interest (KOIs), all of which have been investigated already by other means. I show in the following subsections that this analysis can both easily validate known planets (using no or very minimal follow-up observations) and identify probable false positives. §6.4.1 illustrates the procedure in detail using the example of Kepler-19b, §6.4.2 describes the results of analyzing 17 additional known *Kepler* planets, and §6.4.3 discusses application to 13 known or suspected false positives.

### 6.4.1 Kepler-19b

Kepler-19b, or KOI 84.01, orbits a Sun-like star with a 3.5-day period and has a radius of  $2.2 R_{\oplus}$ . The discovery paper (Ballard et al. 2011) details the sequential battery of tests used to rule out false positive explanations for its photometric signal: first the preliminary tests possible with just the photometry (search for

a secondary eclipse and measurement of the in-transit photocenter shift), and then the follow-up observations. The observation follow-up effort included reconnaissance spectroscopy at three epochs using the McDonald Observatory 2.7m telescope, high-precision radial-velocity Keck/HIRES spectroscopy at 20 epochs (though only 8 were used in analysis), adaptive optics (AO) imaging at the Palomar 200-in, speckle imaging at the WIYN telescope in two filters, and 16 hours of *Warm Spitzer* observation to measure the color dependence of the transit. None of these follow-up measurements revealed any sign of a false positive, and BLENDER analysis combined all these constraints with light curve modeling and estimates of the planet and false positive priors (based on *Kepler* data itself) to calculate a final false positive probability of about 1 in 7000, validating the signal.

For my analysis of Kepler-19b, I begin from the assumption that no secondary eclipse is detected in the photometry (at a level of 200ppm), and that photocenter analysis constrains any potential blend to be within  $2''$  at most. I then proceed through the steps detailed in §6.3: use the target star’s position and KIC photometry to simulate the representative populations for each scenario considered, calculate the priors based on these simulations and various assumptions about stellar populations, measure the trapezoidal shape parameters of the Kepler-19b signal itself, and calculate the likelihoods for each scenario based on these measured shape parameters and the distributions of shape parameters of the scenario. Figures 6.4 and 6.5 illustrate how the shape of the signal matches the typical shape of a transiting planet much better than any of the false positive scenarios. Assuming a 1% specific occurrence rate for planets like Kepler-19b, these calculations result in a FPP of 2.8%. Alternatively, a specific occurrence rate of about 6% would be required for FPP=0.5%—close to being securely validated. The most likely false positive is the HEB scenario (Figure 6.4(c)).

When generating the representative population for the true transiting planet scenario for this signal, I use the stellar parameters for KOI 84 provided in Batalha et al. (2012) ( $0.86 R_{\odot}$ ,  $0.91 M_{\odot}$ ), with 20% uncertainties. These notably do not agree exactly with parameters published in the discovery paper, determined from spectroscopic analysis ( $1.1 R_{\odot}$ ,  $0.97 M_{\odot}$ ). The effect of this discrepancy is visible in Figure 6.5, where the measured slope parameter ( $T/\tau$ ) of the signal appears to fall slightly above the main distribution of the simulated population: a smaller star will result in a smaller  $T/\tau$  for a fixed value of  $R_p/R_*$ . The result is that the likelihood of the planet signal is calculated to be lower than what it would be if the stellar parameters of the simulated population were correct, leading to an *overestimate* of the FPP. Since accurate stellar parameters will not necessarily always be available, this circumstance makes for a realistic test case.

The next step in my re-analysis of the Kepler-19b FPP is to pretend that a single follow-up observation is available: a high-resolution spectrum from which  $T_{\text{eff}}$  and  $\log g$  of the host star can be accurately measured, to a precision of 80 K and 0.1 dex, respectively. The effect of this constraint is to rule out those binary and hierarchical scenarios from the population simulations that were initially allowed (since the overall colors matched the KIC colors), but whose temperatures or surface gravities are  $>3\sigma$  discrepant from this measurement. Figure 6.6 shows how the distribution of shape parameters for the HEB scenario changes under this constraint such that the measured Kepler-19b parameters are even more on the outskirts—the HEB instances

Table 6.1. Test sample: Non-TTV Confirmed Kepler planets<sup>1</sup>

Name	Method	$f_{p,V}$	$f_{p,V}^{\text{spec}}$	$f_{p,V}^{\text{AO}}$	$f_{p,V}^{\text{spec,AO}}$
Kepler-4b	RV	<b>0.006</b>	<b>0.013</b>	<b>0.003</b>	<b>0.010</b>
Kepler-6b	RV	0.161	0.163	0.162	0.163
Kepler-7b	RV	0.068	0.140	0.052	0.112
Kepler-8b	RV	1.227	1.087	0.808	0.781
Kepler-10b	RV/BLENDER	<b>0.014</b>	<b>0.014</b>	<b>0.000</b>	<b>0.000</b>
Kepler-10c	BLENDER	<b>0.014</b>	<b>0.012</b>	<b>0.002</b>	<b>0.000</b>
Kepler-11g	BLENDER	0.289	0.200	0.140	0.061
Kepler-12b	RV	0.086	0.018	0.086	0.018
Kepler-14b	RV	0.010	<b>0.000</b>	<b>0.000</b>	<b>0.000</b>
Kepler-15b	RV	0.293	0.116	0.193	0.104
Kepler-17b	RV	0.109	0.131	0.094	0.118
Kepler-18b	BLENDER	0.460	0.241	0.189	<b>0.002</b>
Kepler-19b	BLENDER	0.058	<b>0.018</b>	<b>0.030</b>	<b>0.001</b>
Kepler-20e	BLENDER	0.071	0.056	<b>0.014</b>	<b>0.001</b>
Kepler-20f	BLENDER	0.178	0.176	<b>0.009</b>	<b>0.007</b>
Kepler-21b	BLENDER	<b>0.010</b>	<b>0.010</b>	<b>0.001</b>	<b>0.001</b>
Kepler-22b	BLENDER	0.048	0.051	<b>0.004</b>	<b>0.001</b>
<b>TOTALS</b>	<b>17 (all)</b>	<b>4</b>	<b>6</b>	<b>9</b>	<b>10</b>

<sup>1</sup>The results of applying this validation procedure to a sample of known *Kepler* planets, 9 of which were confirmed with RV measurements and 9 by BLENDER analysis (one by both).  $f_{p,V}$  is the specific planet occurrence rate required to result in an FPP of 0.5% (Equation 6.8). When  $f_{p,V}$  is significantly below a reasonable occurrence rate estimate, (Figure 6.2), I consider the planet validated (marked in bold). The ‘spec’ and ‘AO’ superscripts indicate the effect of incorporating spectroscopic and imaging follow-up observations. Most of the planets are validated with both types of follow-up; the ones that are not are nearly all giant planets.

that could mimic a transit shape were exactly the ones ruled out by the stellar characterization. Using the likelihoods and priors updated for this constraint, the specific occurrence rate assumption necessary to result in an FPP of 0.5% ( $f_{p,V}$ ) for Kepler-19b is only 1.8%—small enough to result in a secure validation.

To investigate whether a single AO observation would have a similar effect, I remove the spectroscopic constraint and submit the simulated populations to the constraint of a generic deep AO image ( $\Delta K = 8$  at  $0''.7$ ), ruling out all instances of the false positive scenarios that would be identified by this observation. This AO constraint alone also decreases the FPP enough to claim validation, though not quite as strongly as the spectroscopic constraint:  $f_{p,V} = 0.03$ . The combination of the two constraints results in  $f_{p,V} = 0.1\%$ .

The conclusion of this investigation is striking: secure validation of the Kepler-19b signal with these methods would have been possible using either of two single-epoch follow-up observations: a high-resolution spectrum to measure the spectroscopic parameters or an AO image to constrain blends. With both of these observations, validation would be secure beyond doubt.

### 6.4.2 Kepler planet test sample

While Kepler-19b is an eye-opening individual example of the potential of this analysis to fully validate transiting planet candidates with very modest follow-up investment, it is just a single example. I thus identify a broader sample to test in a similar manner: 17 other known *Kepler* planets, selected to be all of the officially designated *Kepler* planets that do not show significant transit timing variations. 9 of these were confirmed with RV measurements, and the additional 8 were BLENDER-validated.

For each of these transit signals I follow the exact same analysis sequence as described above for Kepler-19b, using the same generic assumptions of a 200ppm secondary eclipse limit and a blend radius of  $2''$ . I find that 4 of the 17 would have been validated with no follow-up observations at all, 6 require only stellar characterization to validate, 9 require only AO imaging, and all except 7 (all giant planets but 1) can be validated with the combination of both constraints. These results are summarized in Table 6.1, which presents the values of  $f_{p,v}$  for each signal under the various constraints. A signal is considered validated when  $f_{p,v}$  is significantly below the expected planet occurrence rate for a planet of that size—the values low enough for validation are bolded in the table. These results are quite encouraging, suggesting that perhaps many future *Kepler* planet validations will require only minimal follow-up observations.

### 6.4.3 Known false positives

Despite the encouragement of the above results, it is fair to ask whether this analysis is skewed such that it *always* returns a low false positive probability, and thus automatically tuned for easy planet validation. To address this concern, I investigate a sample of *Kepler* candidates known to be false positives.

KOI 552.01 (Figure 6.7) is a Jupiter-sized candidate whose host star has a KIC-estimated  $T_{\text{eff}} = 6018$  K. This was selected as one of the candidates for follow-up with the SOPHIE spectrograph, and Bouchy et al. (2011) measured it to have  $\sim$ km/s RV variation between two epochs. This suggests that the signal is likely caused by an eclipsing M-dwarf star rather than a giant planet. In addition, analysis of the spectra suggests that the temperature of the star is 6500 K, significantly hotter than the KIC estimate.

For KOI 552.01 I calculate FPP=0.90 (assuming a specific occurrence of 1%, appropriate for Jupiter-sized planets (Wright et al. 2012))—in other words, this procedure clearly identifies this signal as a very likely false positive using only the photometry, and no follow-up information at all. Figure 6.8 illustrates how well the shape of the signal is described by a hierarchical system with an eclipsing binary. Conveniently, an HEB explanation for KOI 552.01 explains not only the eclipse signal but also the discrepancy between the KIC and spectroscopic  $T_{\text{eff}}$  estimates: if this system is indeed a triple system with a primary with  $T_{\text{eff}} = 6500$  K and two cooler stars, then trying to interpret the broadband photometry of the whole system as a single star would naturally result in a cooler temperature estimate.

Similar analysis of 12 other KOIs that have been identified as false positives gives similar results, with all but one having FPP > 20%, and six having FPP > 80% (Table 6.2). Of these, 11 were identified as

Table 6.2. Test sample: False positives<sup>1</sup>

KOI	FP identification method	$\delta(\%)^2$	$T/\tau^3$	FPP <sup>4</sup>
552.01	RV variation <sup>a</sup>	0.8	2.6	0.95
1187.01	color-dependent transit <sup>b</sup>	0.2	2.4	0.82
190.01	Blended CCF; RVs <sup>c</sup>	1.1	3.4	0.22
340.01	RVs <sup>c</sup>	2.1	5.5	0.30
418.01	Blended CCF; RVs <sup>c</sup>	1.2	3.4	0.85
419.01	RVs; sec. eclipse only <sup>c</sup>	0.8	2.4	0.96
425.01	Blended CCF; RVs <sup>c</sup>	1.2	2.8	0.61
607.01	RVs <sup>c</sup>	0.7	3.7	0.32
609.01	double-lined binary <sup>c</sup>	0.4	2.5	0.96
667.01	blend (DSS photometry) <sup>c</sup>	0.9	3.8	0.07
698.01	RVs; sec. eclipse only <sup>c</sup>	0.7	2.5	0.96
1786.01	RVs <sup>c</sup>	0.8	5.8	0.48

<sup>1</sup>The results of applying this analysis on a sample of KOIs that have been identified as false positives with various follow-up observations. Nearly all are identified as likely false positives.

<sup>2</sup>Transit depth

<sup>3</sup>The ratio of the total transit duration to the ingress/egress duration. A value close to 2 is a V-shaped signal; a larger value indicates a more box-shaped transit.

<sup>4</sup>False positive probability, assuming a specific occurrence rate of 1%.

<sup>a</sup>Bouchy et al. (2011)

<sup>b</sup>Colón et al. (2012)

<sup>c</sup>Santerne et al. (2012)

false positives by Santerne et al. (2012) and one (1187.01) was identified by Colón et al. (2012) to have a color-dependent transit depth. I assume a specific occurrence rate of 1% for all these candidates. All but 1187.01 have transit depth  $\geq 0.4\%$ ; this is due to the way that the SOPHIE follow-up sample was chosen. Notably, many of these signals have  $T/\tau$  parameter very close to 2, indicative of a V-shaped transit shape. This demonstrates that incorporating transit shape into this analysis allows this procedure not only to validate planets, but also to reliably identify likely false positives.

#### 6.4.4 Testing assumptions and simulation variance

As mentioned in §6.3.1, the false positive calculations that this procedure delivers are dependent on a set of astrophysical assumptions. In addition, apart from *a priori* assumptions, the calculations are subject to intrinsic variance due to the fact that FP distributions used to calculate the likelihoods are the result of Monte Carlo simulations. To test the degree to which qualitative results depend on these effects, we repeat the calculations for the known planet sample many times, using different sets of assumptions. I repeat each test three times, with the full range of results for the final  $f_{p,V}$  numbers illustrated for each test in Figure 6.9. The tests are the following:

- The original simulations under the original assumptions (repeated three times).
- The original simulations under the original assumptions, but using  $N = 100,000$  for the simulations (rather than  $N = 20,000$ )
- $N = 100,000$ , changing the mass ratio distribution to be flat between 0.2 and 1 (rather than between 0.1 and 1)
- $N = 100,000$ , using the Dartmouth stellar models (Dotter et al. 2008) in place of the Padova models.
- $N = 100,000$ , increasing the “color tolerance” of the EB and HEB scenarios to 0.2; i.e. allowing all FP scenarios that match within 0.2 mag of the  $g-r$  and  $J-K$  colors, rather than 0.1 mag.

As seen in the figure, in only two cases do the range of results cross the nominal  $f_{p,v}$  “validation threshold”: Kepler-11g occasionally drops into validation territory, as does Kepler-7b, for one instance of the  $N = 20,000$  simulation. None of the planets validated with both follow-up observations from Table 6.1 cross over into the non-validation regime. We conclude that larger simulations are always better if possible, to reduce variance, but that  $N = 20,000$  is usually sufficient. While the exact value of FPP may change by a factor of a few from repetition to repetition, secure validations remain secure validations.

There are other input assumptions not tested by this exercise, but these are mostly trivial assumptions that can be tested on a case-by-case basis simply by adjusting the multiplicative factors that go into the prior calculation for a given scenario. These include stellar binary and triple fractions, the fraction of binaries that are in “short” orbits, the occurrence rate of planets around background stars, and the density of background stars returned by TRILEGAL. Adjusting any of these numbers changes the prior factor for that scenario, and correspondingly changes the false positive probability. If the FPP is small and one false positive scenario dominates, the overall FPP is directly proportional to the prior factor for that scenario (see the discussion in §5 of MJ11); e.g., if the HEB scenario is the most likely false positive for a particular candidate, doubling the assumed stellar triple fraction will approximately double the FPP (or, equivalently, double the  $f_{p,v}$  required for validation).

## 6.5 Relation to Morton & Johnson (2011)

While I have demonstrated that the results from this newly updated FPP analysis are reliable both with known planets and known false positives, I now briefly discuss why these results may differ from the individual FPPs published in Morton & Johnson (2011a). First of all, it is important to understand that the goal of the previous work was to demonstrate that, in general, *Kepler* candidates *that pass all possible initial photometric vetting* are very likely to be planets. That is, the MJ11 analysis is only strictly valid for signals that are not obviously V-shaped and whose light curves have been declared free from secondary eclipses.

As it turns out, many of the candidates from Borucki et al. (2011) were not in fact fully vetted. This was made clear in that paper, but this fact was not reflected in the Morton & Johnson (2011a) KOI FPP table. In other words, there are some signals that are clearly V-shaped (a number of these are now in the known false positive sample discussed in §6.4.3), and likely also signals that show faintly detectable secondary eclipses. For such KOIs, it is probable that the final FPP will turn out to be significantly higher, as is the case for the sample of false positives detected by Santerne et al. (2012).

The analysis of Morton & Johnson (2011a) was neither a planet validation procedure nor a false positive detection procedure, nor was its purpose to claim that the *Kepler* results are completely immune from the problem of false positives. Instead, its goal was to demonstrate that the *a priori* FPP for fully vetted *Kepler* candidates is typically low, thus suggesting that the overall sample of candidates generally reflects the properties of the true sample of planets (barring the existence of a pathologically large number of V-shaped signals among the candidate list). On the other hand, this work, having advanced the techniques of Morton & Johnson (2011a), is intended to validate candidates, using all available information on an individual basis.

## 6.6 Conclusions

Transiting exoplanet surveys continue to grow much faster than the capabilities of traditional follow-up infrastructure. While the observations required to positively confirm candidates have always been the bottleneck for transit surveys, the *Kepler* mission has made this issue much more acute. Future ground- and space-based surveys will doubtless continue to produce orders of magnitude more candidates than can be securely confirmed either with radial velocity or transit timing variation measurements. As a result, the onus of confirmation for the majority of transiting planet candidates has shifted irrevocably from dynamical confirmation to probabilistic validation.

With probabilistic validation playing such a central role, the procedure I describe in this paper enables a revolution in how transit surveys could operate: *validating more planets, using fewer follow-up resources*. Recall what I have demonstrated about this procedure:

- It is efficient: an end-to-end FPP calculation for a transit signal takes only about 10 minutes as currently implemented—and there is room for further computational optimization.
- It often requires only one or two single-epoch follow-up observations (either spectral characterization, AO imaging, or both—and sometimes neither) to validate a true transiting planet.
- It can reliably identify false positives.

The benefits of applying this analysis to a large sample of candidates from a transit survey such as *Kepler* are evident. Spectroscopic follow-up can be prioritized in order of increasing FPP, likely leading directly to many validations. Preference for AO imaging can then be given to those candidates not validated with a spectrum. RV observations of likely false positives can be avoided, and more time devoted to RV measurements of likely



planets or characterization spectroscopy. And the most resource-intensive follow-up tools, such as multi-band photometry, validation-inspired RV measurements, and the full BLENDER analysis, can be reserved for the most stubborn candidates—those that are not likely false positives, but resist validation even after spectral characterization and AO imaging. A good example of a candidate ideal for this last-stage analysis would be one with a residual FPP of 5-10% or so, for which AO imaging reveals a faint blended companion.

With regard to *Kepler*, I intend to apply this procedure to *all* of the remaining unvalidated KOIs, using whatever follow-up observations are available. There are several ongoing projects to spectroscopically characterize and obtain AO images of large numbers of KOIs; doubtless these efforts have already provided the measurements necessary to validate hundreds of planets.

The ultimate goal of this project is to make self-contained, stand-alone transiting planet validation software available to the community. This will allow present and future transit surveys to apply this analysis at an early stage, hopefully enabling follow-up resources to be optimized and planet validations to be expedited. The hope is that with fewer observational resources dedicated to the task of planet validation, more can be devoted to detailed characterization of individual systems of particular interest. In addition, with large numbers of securely validated transiting planets, statistical population studies will become possible without worrying about any residual effects from false positive contamination.

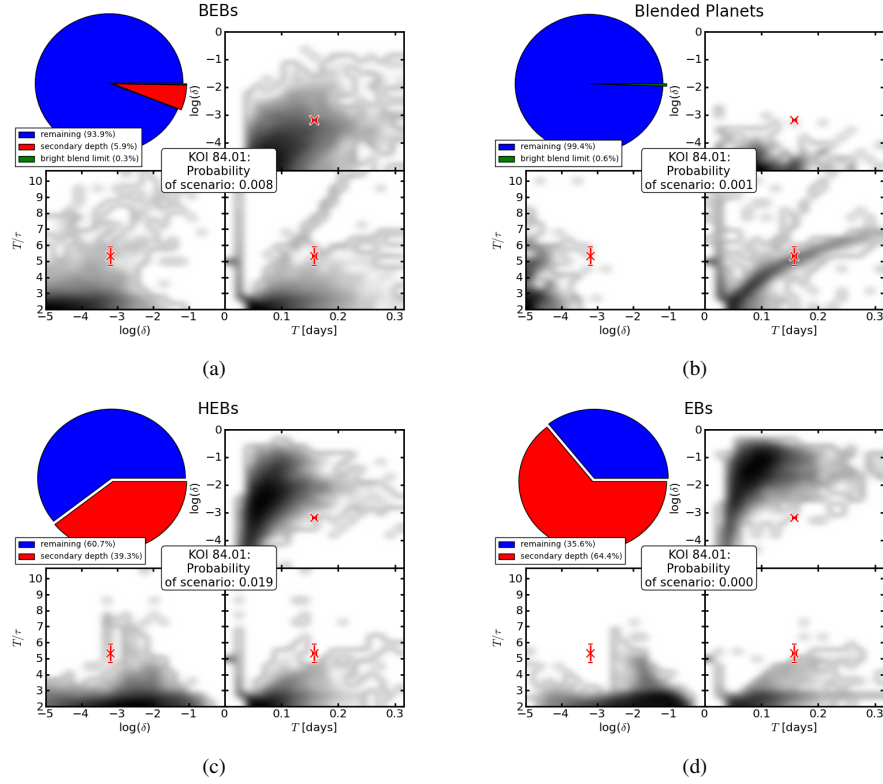


Figure 6.4 The false positive landscape for Kepler-19b (KOI 84.01). Each of these plots illustrates a three-dimensional probability distribution for the trapezoidal shape parameters (depth  $\delta$ , duration  $T$ , and “slope”  $T/\tau$ ) for a false positive scenario (see §6.2 for descriptions of the scenarios). Each of these distributions is made by simulating a statistically representative population (§6.3.1) for a scenario and fitting the shape parameters to each simulated instance of the scenario. Each population begins with 20,000 simulated instances, and only instances that pass all available observational constraints are included in these distributions—in this case the constraints are that the blended star be at least 1 mag fainter than the primary target, the lack of an observed secondary eclipse deeper than 200ppm, and that the KIC  $g-r$  and 2MASS  $J-K$  colors match within 0.1 mag. The pie chart for each scenario illustrates what fraction of the initial simulations pass these tests. On each plot the shape parameters of the transit signal are marked, with the ‘X’ showing the median of an MCMC fit, and error bars illustrating 95% confidence. The probability of each scenario is calculated by dividing the prior  $\times$  likelihood for that particular scenario by the sum of prior  $\times$  likelihood for all the scenarios (including the planet scenario illustrated in Figure 6.5). Priors for each scenario are calculated according to §6.3.2, and the likelihoods are calculated by integrating the illustrated distributions over the observed measurement.

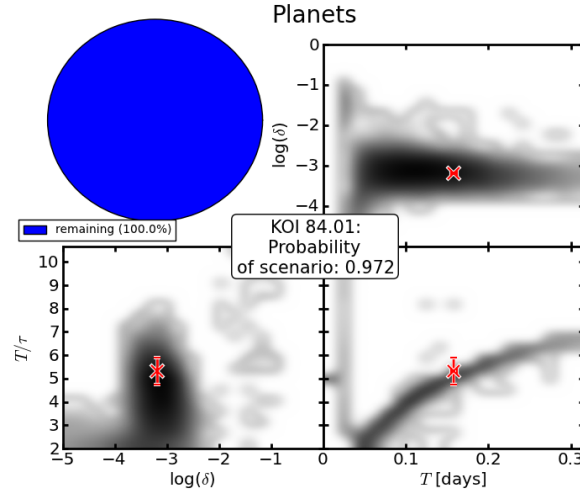


Figure 6.5 The transiting planet scenario for Kepler-19b (KOI 84.01). The plots illustrate the three-dimensional probability distribution for the trapezoidal shape parameters for simulations of a representative population (§6.3.1) of transiting planets around KOI 84. The radius distribution of planets in this population is chosen to be a narrow range centered on the radius derived for KOI 84.01, assuming it is a *bona fide* transiting planet. The spread and shape of this distribution are caused by different-sized planets in the simulation, different stellar properties (according to uncertainties in stellar radius and mass), and by variations in orbital inclination, which changes duration and slope in a correlated manner (bottom-right panel). The assumed occurrence rate of planets in this radius bin for this calculation is 1% (see §6.3.4 for more discussion about this “specific occurrence rate”). The probability of the Kepler-19b signal being a transiting planet, given the false positive landscape illustrated in Figure 6.4 (constrained by only the signal photometry and KIC colors) is about 97%, giving  $\text{FPP} = 0.03$ . The signal does not fall quite exactly in the middle of the  $T/\tau$  distribution because this population simulation uses photometrically estimated physical properties (mass, radius) for the host star, rather than the more accurate spectroscopically derived properties. Note that uncertainties like this will tend to decrease the planet likelihood, and thus will contribute to *overestimating* the FPP.

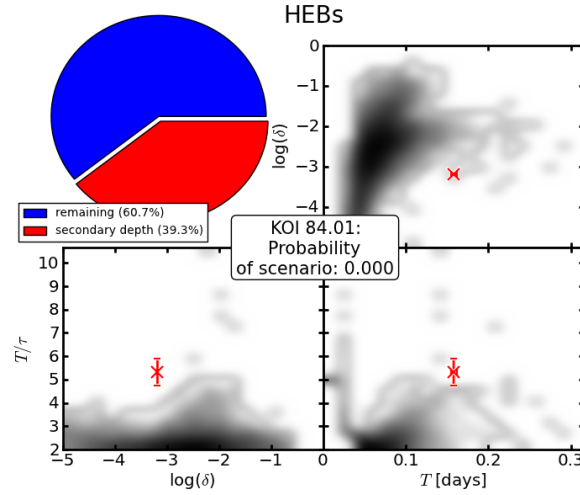


Figure 6.6 The hierarchical triple eclipsing binary (HEB) scenario for Kepler-19b (KOI 84.01) under the constraint of spectroscopic characterization of the host star. The plots illustrate the three-dimensional probability distribution for the trapezoidal shape parameters for simulations of a representative population ( $\approx 3.1$ ) of HEB false positive scenarios, but only including those instances in which the spectroscopic properties ( $T_{\text{eff}}$ ,  $\log g$ ) of the primary star match those that were measured for KOI 84. Compare this distribution to Figure 6.4(c): the systems whose shape parameters overlapped with the properties of the measured signal are not allowed by the spectroscopic characterization—these correspond to the higher-mass systems illustrated in Figure 6.1. This constraint allows for the signal to be validated, as the HEB scenario was the most likely FP scenario before including the follow-up spectral information.

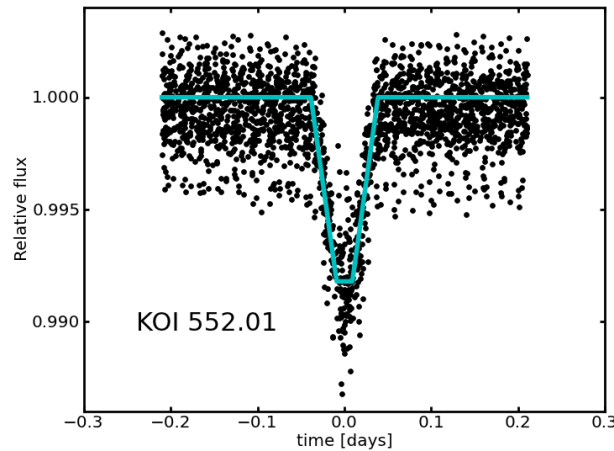


Figure 6.7 The phase-folded *Kepler* photometry of KOI 552.01. The solid line illustrates the best-fitting trapezoid model used for the likelihood calculations (§6.3.3). Bouchy et al. (2011) measured large RV variations of the host star, suggesting that this signal is caused by an eclipsing M-dwarf rather than a planet. Note that the signal is clearly more V-shaped than box-shaped.

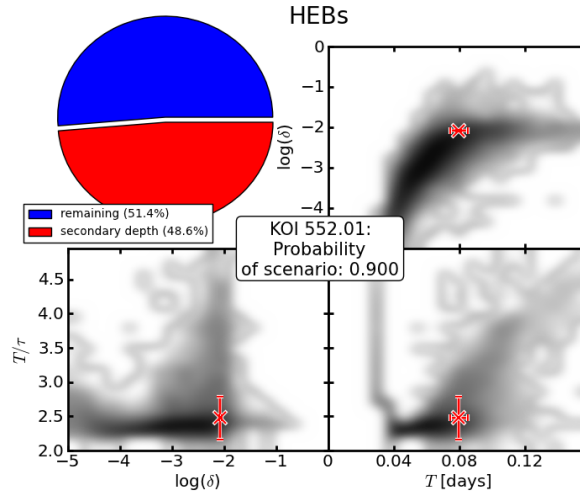


Figure 6.8 The hierarchical triple eclipsing binary system scenario for KOI 552.01. This KOI was measured by Bouchy et al. (2011) to have  $\sim \text{km/s}$  radial velocity variation between two epochs, suggesting a stellar eclipse rather than a transiting planet—an observed astrophysical false positive. The analysis presented in this paper gives an FPP of  $>90\%$  for this signal, and this figure illustrates why: the shape of the signal is exactly consistent with what would be expected from a hierarchical triple system including an eclipsing binary.

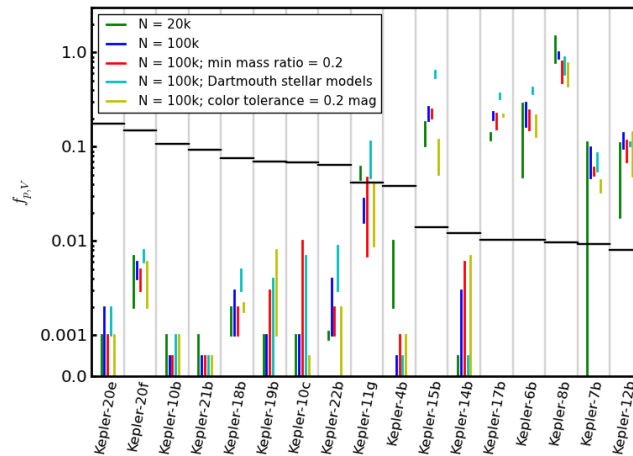


Figure 6.9 The results of testing the effects of sample variance and changing various assumptions on the false positive calculations described in this paper. Each colored vertical line represents the full range of  $f_{p,v}$  (the specific occurrence rate assumption required for validation) over three iterations of the procedure for each planet from Table 6.1, ordered by increasing radius. The five different vertical lines for each planet represent five different experiments, varying the size of the Monte Carlo simulations and different input astrophysical assumptions. The horizontal black lines represent the value of  $f_{p,v}$  inferred from an assumption of an overall 40% planet occurrence rate with  $dN/dR \propto R^{-2}$ . Below the line means the signal is securely validated. All of the planets that were claimed as validated by the original analysis (the final column of Table 6.1) remain validated for each iteration of each of these tests.

## Chapter 7

# The Radius Distribution of Planets Around Cool Stars

*This chapter is the manuscript of a paper that has been submitted to ApJ by TDM and Jonathan Swift and is currently under review.*

### 7.1 Introduction

The discovery of the first exoplanets (Wolszczan & Frail 1992; Mayor & Queloz 1995; Marcy & Butler 1996) has sparked tremendous growth in research and interest in the formation and evolution of planetary systems beyond the Solar System. Not unlike many areas of astronomy, however, the first discoveries are not representative samples; rather, “hot Jupiters” are relatively rare (Wright et al. 2012; Howard et al. 2010) in comparison to the new populations of exoplanets now being revealed by the *Kepler* Mission (Borucki et al. 2011; Batalha et al. 2012; Burke 2013). The most common kinds of planets within *Kepler*’s discovery space of  $R_p \gtrsim 0.5R_\oplus$ , and  $P \lesssim 100$  d appear to be somewhat larger than Earth but smaller than Neptune,  $1 < R_p < 4R_\oplus$  (Howard et al. 2012; Fressin et al. 2013; Dressing & Charbonneau 2013).

Much of our understanding of planet formation is anchored in decades of research into our own solar system. But now the burgeoning exoplanet population provides us with a new context, revealing important insights into planet formation throughout the Galaxy. For example, the large amount of planetary mass seen close to host stars is evidence that protoplanetary disks may have much higher surface densities than previously thought (Hansen & Murray 2012; Chiang & Laughlin 2012) or that the observed planets migrated from regions further from their host star where more mass was readily available for assembly (Swift et al. 2013).

The multi-transit systems of the *Kepler* sample also offer a wealth of information regarding their formation and evolution (Lissauer et al. 2011; Fabrycky et al. 2012b). The period ratios of planets within a given system show a propensity to lie just outside of first order mean motion resonances (Fabrycky et al. 2012a; Steffen et al. 2013), which may be an imprint of dissipative (Lithwick & Wu 2012; Batygin & Mor-

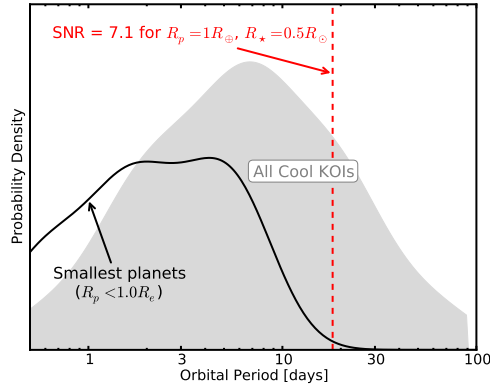


Figure 7.1 Evidence supporting the hypothesis that small planets are incomplete in the Cool KOI sample. The solid black line is the observed (smoothed) distribution of planets smaller than  $1 R_{\oplus}$ ; the grey shaded area is the observed period distribution of all the Cool KOIs. Neither distribution is corrected for transit probability. The vertical dashed red line indicates the period at which a  $1 R_{\oplus}$  planet around a  $0.5 R_{\odot}$  star (typical of the Cool KOI sample) would have SNR of 7.1, the nominal detection threshold for KOI identification. The lack of observed small planets at periods longer than 10 days is thus very plausibly due to incompleteness.

bidelli 2013) or stochastic mechanisms (Rein 2012) in the formation or evolution of planetary systems. The low inferred mutual inclination of multi-transit systems ( $\sim 1^{\circ}$ – $3^{\circ}$ ; Fabrycky et al. 2012b; Fang & Margot 2012) together with the relative number of single versus multi-transit systems provides constraints on the number of planets in a given system within *Kepler*’s discovery window, else it may be the first indication of a separate, high-inclination population of single transit systems (Hansen & Murray 2013; Fang & Margot 2012). Lastly, the mutual gravitational interactions within some multi-planet systems offer an estimation of the planet masses (Lithwick et al. 2012) that then inform planetary compositions and atmospheric evolution scenarios (Rogers et al. 2011; Wu & Lithwick 2012; Lopez et al. 2012).

In this article we focus on yet another important clue regarding the formation of the compact systems revealed by *Kepler*: the distribution of planetary radii. The initial estimates of the planet radius distribution by Howard et al. (2012) showed a dramatic increase in the number of planets at ever smaller size. Citing incompleteness, however, they did not follow this trend in their analysis to planet radii smaller than  $2 R_{\oplus}$ . In an independent study by Youdin (2011), a parametric estimation of the planetary distribution function revealed a deficit of large planets in short period orbits that would support a core accretion then migration formation scenario.

More recent estimates of the planet radius distribution show a preferential size scale in the *Kepler* sample indicated by a flattening and possible turnover in the log-binned histogram of detected planet candidates somewhere around  $2 R_{\oplus}$  (Fressin et al. 2013; Dressing & Charbonneau 2013; Petigura & Marcy 2013). If true, this would be an important clue toward understanding the key mechanisms that shape the observed population of compact planetary systems that pervade the Galaxy. However, these analyses are constrained by the limitations of coarse histograms; no analysis to date has yet characterized the shape of the exoplanet radius

distribution in enough detail to allow meaningful comparison to planet formation and evolution theories.

With this in our sights, we focus on the smallest stars in the *Kepler* Object of Interest (KOI) sample for two reasons: (1) between the spectroscopic studies of the M dwarf sample by Muirhead et al. (2012b) and the photometric re-calibration of the Kepler Input Catalog (KIC) for the coolest stars by Dressing & Charbonneau (2013), the “Cool KOIs” constitute a well characterized sample, and (2) since a transit signal is proportional to the square of the planet to star radius ratio, the Cool KOIs are optimal for probing the planet radius distribution for the smallest planet sizes.

The particular goal of this work is to derive the shape of the planet radius function *properly marginalized over orbital period*. Figure 7.1 illustrates why this is an issue: the smallest planets are not complete out to the same orbital periods as larger planets; thus careful correction is required in order to achieve this goal. These concepts of incompleteness and period bias were studied in detail early in the history of transit surveys (Pepper et al. 2003; Gould et al. 2003; Gaudi 2005), but have yet to be applied in detail to *Kepler* data.

In §7.2 we walk through the steps required to properly extract a non-parametric empirical estimate of the true planet radius function given a population detected in a well-characterized transit survey. In §7.3 we apply these methods to the Cool KOIs to derive the radius distribution for small planets around small stars. We explore the various assumptions that go into this calculation in §7.5, and conclude in §7.6.

## 7.2 Formalism

We define the planet radius distribution function  $\phi_r^{P_{\max}}(r)$  such that

$$\int_{r_{\min}}^{r_{\max}} \phi_r^{P_{\max}}(r) dr = \text{NPPS}, P < P_{\max}; \quad (7.1)$$

that is, a density function with an overall normalization giving the average number of planets per star (NPPS) for planets with period less than  $P_{\max}$  days, for planet radii  $r$  between  $r_{\min}$  and  $r_{\max}$ . The problem of calculating planet occurrence rates from *Kepler* has been quite an industry over the last few years (Youdin 2011; Howard et al. 2012; Dong & Zhu 2012; Swift et al. 2013; Fressin et al. 2013; Petigura & Marcy 2013; Dressing & Charbonneau 2013). However, there has been little quantitative discussion of deriving the detailed shape of the radius function beyond drawing histograms. In the following subsections, we review and refine the general principles of an occurrence calculation and then describe how to follow these principles to construct a non-parametric empirical radius function that obeys the above desired properties.

### 7.2.1 Occurrence Calculations

In a perfectly idealized survey that is both 100% reliable and 100% complete, the occurrence rate of planets is simply

$$\text{NPPS} = \frac{N_p}{N_{\star}}, \quad (7.2)$$



where  $N_p$  is the number of detected planets and  $N_*$  is the number of stars surveyed. In practice, however, this must be corrected for both incompleteness and unreliability as follows:

$$\text{NPPS} = \frac{1}{N_*} \sum_{i=1}^{N_p} w_i. \quad (7.3)$$

Here the sum is over all detections and  $w_i$  is a weighting factor applied individually to account for the various necessary corrections. Generally, these weights can be thought of as

$$w_i = \frac{(1 - \text{FPP}_i)}{\eta_i}, \quad (7.4)$$

where  $\text{FPP}_i$  is the probability that signal  $i$  is a false positive and  $\eta_i$  is an individualized efficiency factor for the detection of planet  $i$ . In this work, we do not incorporate the calculations of  $\text{FPP}_i$  in detail, as the *a priori* false positive rate among candidates is low (Morton & Johnson 2011b), and ongoing analysis (Morton et al., in prep) according to the false positive-calculating procedure of Morton (2012) indicates the false positive rate in this particular sample is negligibly low.

We thus focus on the detection efficiency  $\eta_i$ , which is defined by the following thought experiment: *If a very large number of planets identical to planet  $i$  were distributed randomly around all the stars in the survey, only a fraction  $\eta_i$  could have been detected.* This can be further factored (following Youdin 2011):

$$\eta_i = \eta_{\text{tr},i} \cdot \eta_{\text{disc},i}, \quad (7.5)$$

where  $\eta_{\text{tr}}$  is the geometric transit probability, and  $\eta_{\text{disc}}$  is the “discovery efficiency”: the fraction of planets in this thought experiment *with transiting orbital geometries* that could have been detected by the survey. In previous *Kepler* occurrence rate calculations (Howard et al. 2012; Swift et al. 2013; Dressing & Charbonneau 2013), this factor has been defined as

$$\eta_{\text{disc},i} = \frac{N_{*,i}}{N_*} \quad (7.6)$$

where  $N_{*,i}$  is the number of target stars around which planet  $i$  could have been detected. As we show, however, this is not sufficient to properly characterize  $\eta_{\text{disc},i}$ ; more care must be taken.

Central to correctly calculating  $\eta_{\text{disc}}$  is the fact that whatever transit detection pipeline is used, it is incomplete, especially near the detection threshold. This pipeline incompleteness *must* be carefully considered in any occurrence rate calculation. The analysis of Petigura & Marcy (2013) is a model of one way this can be done: simulating planets throughout the radius and period parameter space considered in order to directly measure  $\eta_{\text{disc}}$  as a function of planet radius and period.

A more general conceptual way to attack this problem—and the only one available if one is relying on the results of someone else’s detection pipeline—is to assume that the detection efficiency of any pipeline is a function only of the signal-to-noise ratio (SNR) of the transit signal. This was the approach taken by Fressin

et al. (2013), who determined that for the Batalha et al. (2012) Q1-Q6 catalog, the detection efficiency of the *Kepler* pipeline could be modeled by a continuous “SNR ramp” function of the following form:

$$\begin{aligned}\eta_{\text{SNR}}(\text{SNR}) &= 0; & \text{SNR} \leq S_0 \\ &= \text{linear}; & S_0 < \text{SNR} < S_1 \\ &= 1; & \text{SNR} \geq S_1,\end{aligned}\tag{7.7}$$

where in Batalha et al. (2012)  $S_0 = 6$  and  $S_1 = 16$ . This is notably different from a sharp detection threshold at  $\text{SNR} = 7.1$ , which was used by both Swift et al. (2013) and Dressing & Charbonneau (2013) in their occurrence calculations. The newly released and more uniformly vetted Q1-Q8 KOI catalog, currently hosted at the NASA Exoplanet Archive, is better characterized by a steeper SNR ramp (F. Fressin, priv. comm.), as it is more complete than Batalha et al. (2012); thus, we adopt an SNR ramp where  $S_0 = 6$  and  $S_1 = 12$ .

With this function defined,  $\eta_{\text{disc},i}$  can then be calculated by the following procedure: simulate planet  $i$  around every star in the survey, each of which has both a different radius and different photometric noise properties, to obtain a (normalized) distribution of SNRs  $\phi_{\text{SNR},i}$  for that planet, and then marginalize the detection efficiency over that distribution:

$$\eta_{\text{disc},i} = \int_0^\infty \eta_{\text{SNR}}(s) \cdot \phi_{\text{SNR},i}(s) ds.\tag{7.8}$$

An extremely important consideration in this procedure of constructing  $\phi_{\text{SNR},i}$  is properly treating orbital period. Remember, the ultimate goal of this analysis is to calculate the distribution of planet radii with period less than  $P_{\text{max}}$ :  $\phi_r^{P_{\text{max}}}$ . Thus, in simulating the population of planet  $i$  clones around other stars, it is important to assign each of these clones an orbital period  $P < P_{\text{max}}$  according to a reasonable estimate of the true planet period distribution  $\phi_P$ .

Distributing the hypothetical planets according to a period distribution is crucial because both SNR and planet occurrence are functions of orbital period. For example, imagine that a survey of 1000 stars detects one  $0.5 R_\oplus$  planet in a 1-day orbit with low SNR (e.g.  $\text{SNR} = 10$ ). An occurrence analysis in the style of Equation 7.6 might conclude that this planet had a 20% transit probability and would have been detectable around only half the stars in the survey, thus giving it a weight factor of  $w_i = 10$  and leading to the conclusion that  $0.5 R_\oplus$  planets are rare, only existing around 1% of stars. A slightly more sophisticated analysis might note that  $\eta_{\text{SNR}}(10) = 0.4$ , and give another factor of 2.5 boost, concluding that planets of this size exist around only 2.5% of stars.

However, this conclusion would still be incorrect, since it does not account for the fact that planets of this size may only be detectable at very short periods. What if only a very small fraction of all planets happen to have periods as short as 1 day? *In this case, the supposed rarity of  $0.5 R_\oplus$  planets would be just*

*a misinterpretation of the fact that planets with 1-day orbits are rare.* The small planets that doubtless still do exist at larger orbital periods have not been detected, and no correction has been made to account for this. When constructing  $\phi_{\text{SNR},i}$ , distributing the hypothetical planets according to a reasonable period distribution will avoid this misdiagnosis. This was not done in either Swift et al. (2013) or Dressing & Charbonneau (2013), leading both analyses to underestimate the occurrence rate of small planets around small stars. In §7.3 we show this also makes a qualitative difference in the interpretation of the planet radius function.

## 7.2.2 Estimating the Radius Distribution Function

In all the *Kepler* planet occurrence calculations to date, the shape of the radius function has been explored only very coarsely, by calculating the occurrence rate in several different radius bins and either fitting a power law or qualitatively commenting on the shape. Howard et al. (2012) found a good fit to an  $R^{-2}$  power law down to  $2 R_{\oplus}$ , and declined to comment for smaller planets. On the other hand, Fressin et al. (2013) and Petigura & Marcy (2013) note that the occurrence rate of planets increases towards smaller radius but then appears to flatten out below about  $2.8 R_{\oplus}$ . Dressing & Charbonneau (2013) claim that the occurrence rate begins to decrease for planets smaller than  $1.4 R_{\oplus}$ .

Investigating the shape of the radius distribution in more detail requires a non-parametric approach, and also should avoid binning. Here we introduce the concept of using a weighted kernel density estimator (wKDE; e.g. Wang & Wang 2007) in order to accomplish this.

A standard kernel density estimator (KDE) attempts to estimate the true underlying probability distribution of a sample of data points using a function of the following form:

$$\hat{\phi}(x) = \frac{1}{N} \sum_{i=1}^N k(x - x_i; \sigma_i), \quad (7.9)$$

where  $N$  is the number of data points and  $k(x)$  is a zero-mean, normalized kernel function of arbitrary shape (commonly a Gaussian), with some width  $\sigma_i$ , that most generally can be different for each data point. This creates a smooth distribution out of a discrete data set, with the degree of smoothness controlled by the width parameter. The choice of width has tradeoffs in both directions: if the kernels are too narrow the estimator will be bumpy, but if they are too wide they can wash out real structure in the distribution. Often the width is selected to be the same for all points based on the number of data points, or sometimes a variable-width kernel is used, e.g. the distance to the  $n$ th nearest neighbor. The  $1/N$  normalization factor assures that the integral of this density estimator over the whole parameter space is unity.

In order to use the KDE concept to properly reconstruct the radius function of planets detected in a transit survey, each data point has to be weighted appropriately, leading to a wKDE:

$$\hat{\phi}_r^{P_{\max}}(r) = \frac{1}{N_{\star}} \sum_{i=1}^{N_p} w_i \cdot k(x - x_i; \sigma_i), \quad (7.10)$$

where  $w_i = 1/\eta_i$  are the appropriately calculated individual weight factors that renormalize the kernels to correct for missing planets, as discussed in §7.2.1. The weights ensure that the shape of the radius function responds appropriately to the individual corrections, and the  $N_*$  overall normalization ensures that the integral over all radii will return the NPPS, as desired in Equation 7.1. A very natural choice for the  $\sigma_i$  in this case, which avoids having to choose an arbitrary smoothing factor, is the uncertainty in each planet’s radius, most of which comes from uncertainty in the radius of the host star. If this does not make for a sufficiently smooth distribution, then the  $\sigma_i$  can be multiplied by an additional factor to increase the smoothing.

### 7.3 Calculating the Cool KOI Radius Function

One of the biggest concerns to date about interpreting *Kepler* data is uncertainty about stellar parameters. This applies both because the properties of the transit host stars are unknown (derived planet radius depends directly on the radius of the host star) and because the properties of the stars in the survey parent sample are unknown (i.e. is *Kepler* actually surveying dwarf stars or is the parent sample significantly contaminated by giants or subgiants? (Mann et al. 2012)).

Focusing on *Kepler* candidates around relatively low-mass stars alleviates these concerns. Many of these stars have spectroscopically measured stellar properties (Muirhead et al. 2012a; Mann et al. 2012), and in addition, the properties of the parent sample of target stars have been carefully characterized photometrically by Dressing & Charbonneau (2013). Such an investigation thus is narrower than attempting to use the whole *Kepler* sample, but the assurance of a good understanding of the stellar parameters of both the host stars and the general survey sample more than compensates for this loss of generality. In addition, focusing on these “Cool KOIs” enables detailed study of the radius distribution of Earth-sized and smaller planets.

To construct the planet radius function, we thus select the 113 planet candidates with periods  $<90$ d identified in the cumulative KOI catalog posted at the NASA Exoplanet Archive that are hosted by stars with  $T_{\text{eff}} < 4000$ K as characterized by Dressing & Charbonneau (2013). To this sample we add the three KOI-961/Kepler-42 planets, which were left out of the Dressing & Charbonneau (2013) sample because its broad-band colors are consistent with classification as either a giant or a dwarf, even though it has been spectroscopically confirmed to be a  $\sim 0.15 M_{\odot}$  dwarf (Muirhead et al. 2012b). For stellar parameters we use the results presented in Dressing & Charbonneau (2013), except for those KOI host stars that have been spectroscopically characterized according to the observations and procedures described in Muirhead et al. (2012b), for which we use the spectroscopic parameters. As this spectroscopic method is known to be unreliable for  $T_{\text{eff}} > 3800$  K, we defer to the Dressing & Charbonneau (2013) parameters for stars in this temperature range.

In the following subsections, we describe the steps necessary to calculate  $\hat{\phi}_r^{90}$ , the estimate of the radius function for planets on orbits  $<90$ d, from this KOI sample. As described in §7.2, the crucial step toward properly estimating the radius function is calculating the weight factor  $w_i = 1/\eta_i$  for each detection, which includes a transit probability factor and a completeness factor  $\eta_{\text{disc},i}$  (Equation 7.8). Key to calculating  $\eta_{\text{disc},i}$

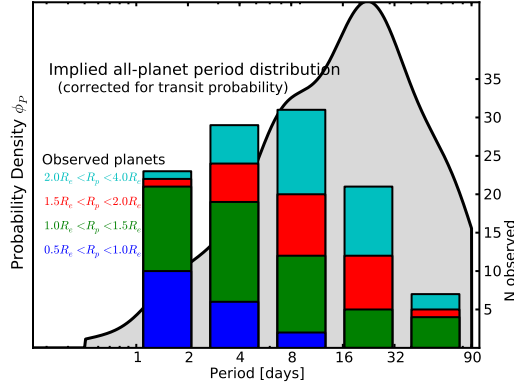


Figure 7.2 The period distribution of planets around *Kepler*'s M dwarfs. The grey shaded region is the implied period distribution of all planets combined, correcting for the effects of transit probability. The bar charts show the observed numbers of planets of different sizes in each period bin. Note the declining fraction of small planets as a function of period—this is most likely an effect of declining detection efficiency for smaller planets on longer-period orbits, and this must be properly accounted for when constructing the planet radius function. The radius function calculation in this paper assumes that all planets are distributed according to the shaded distribution, regardless of planet radius. See §7.5 for a discussion of this assumption.

is determining the SNR distribution of a hypothetical population of clones of planet  $i$  around all the target stars, or  $\phi_{\text{SNR},i}$ , which in turn requires an assumption of the intrinsic period distribution of planets  $\phi_P$ .

### 7.3.1 Period distribution

In order to estimate the shape of the true period distribution of planets of all sizes, we make the simplifying assumption that the period distribution of planets is independent of their radii (see §7.5 for a discussion regarding this assumption). We thus construct the distribution of  $\log P$  from all the planet candidates in the sample, using a wKDE as described in §7.2.2. For the weights we use only the inverse transit probabilities, and enforce that the whole distribution is normalized to unity, creating the probability density function for  $\log P$ . For the widths we use  $\sigma = 0.15$  (in  $\log P$ ), to create a smooth distribution. This is the period distribution function  $\phi_P$  that we use in the following subsection, shown as the grey shaded region in Figure 7.2.

### 7.3.2 SNR distribution

The SNR of a transit signal is usually defined as follows:

$$\text{SNR} = \frac{\delta}{\sigma} \sqrt{N_{\text{tr}} \cdot N_{\text{pts}}}, \quad (7.11)$$

where  $\delta$  is the transit depth,  $\sigma$  is the one-point photometric uncertainty,  $N_{\text{tr}}$  is the number of transits observed, and  $N_{\text{pts}}$  is the number of photometric points per transit. It can be shown that, for a fixed planet radius, SNR should scale with host star radius  $R_*$ , orbital period  $P$ , time observed  $T_{\text{obs}}$ , transit duration  $T_{\text{dur}}$ , and  $\sigma$  as

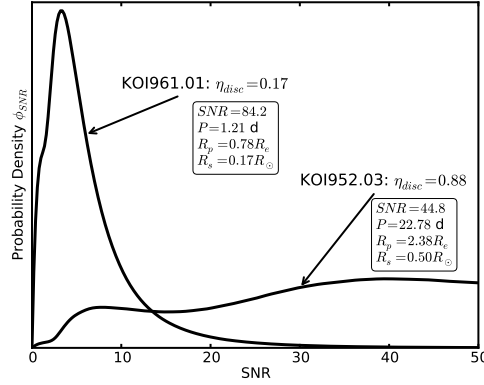


Figure 7.3 Two examples of the SNR distributions resulting from simulating the transit of a given planet around every target star, given randomly assigned periods and impact parameters. The properties of this distribution depend on the properties of the detected system, and the integral of the pipeline detection efficiency function over this distribution gives the “discovery fraction”  $\eta_{\text{disc}}$ . KOI-961.01, a sub-Earth-sized planet in a very short orbit around a very small star (Muirhead et al. 2012b), would have been detectable in only about 1/6 of potential configurations, whereas KOI-952.03, a larger planet around a larger star (Swift et al. 2013), would be detectable in almost any configuration, even though its actual SNR is smaller than that of KOI-961.01.

follows:

$$\text{SNR} \propto R_{\star}^{-2} P^{-\frac{1}{2}} T_{\text{obs}}^{\frac{1}{2}} T_{\text{dur}}^{\frac{1}{2}} \sigma^{-1}, \quad (7.12)$$

where  $T_{\text{dur}}$  itself is a function of  $P$ , scaled semimajor axis  $a/R_{\star}$ , and impact parameter  $b$ . The putative SNR of a planet transplanted from its current configuration to a different period, impact parameter, and host star may thus be calculated by scaling the original SNR appropriately.

For each *Kepler* target star, data on the photometric uncertainty  $\sigma$  is available on a quarter-by-quarter basis, quantified by the “combined differential photometric precision” (CDPP) values on 3-hr, 6-hr, and 12-hr time intervals, which in principle should allow for calculation of the SNR for each transit signal. However, the KOI catalogs also provide SNR for each identified planet candidate, and we find that using CDPP values and Equation 7.11 to calculate SNR does not reliably reproduce the catalog values (it typically underestimates by about 30%, with significant scatter). And since the SNR ramp efficiency characterization (ramp from 0 at SNR = 6 to 1 at SNR = 12) was developed using the catalog SNR values, those are the SNRs we use. We do assume, however, that SNR still scales according to Equation 7.12.

And so, for each of the planet candidates  $i$  in the Cool KOI sample, we take its SNR as provided by the Exoplanet Archive and construct  $\phi_{\text{SNR},i}$  by simulating the entire population of hypothetical alternative configurations: 10,000 iterations of randomly chosen periods (according to the distribution  $\phi_P$  described in §7.3.1) and impact parameters (according to a uniform distribution from 0 to 1) around each star in the target population, calculating the appropriate SNR for each instance according to the scalings in Equation 7.12. As the total number of simulated SNRs is sufficiently large ( $\sim 10^7$ ), the smooth final shape of  $\phi_{\text{SNR},i}$  is defined

by interpolating a histogram with bin width  $\Delta\text{SNR} = 1$ . Figure 7.3 illustrates examples of this distribution for two KOIs.

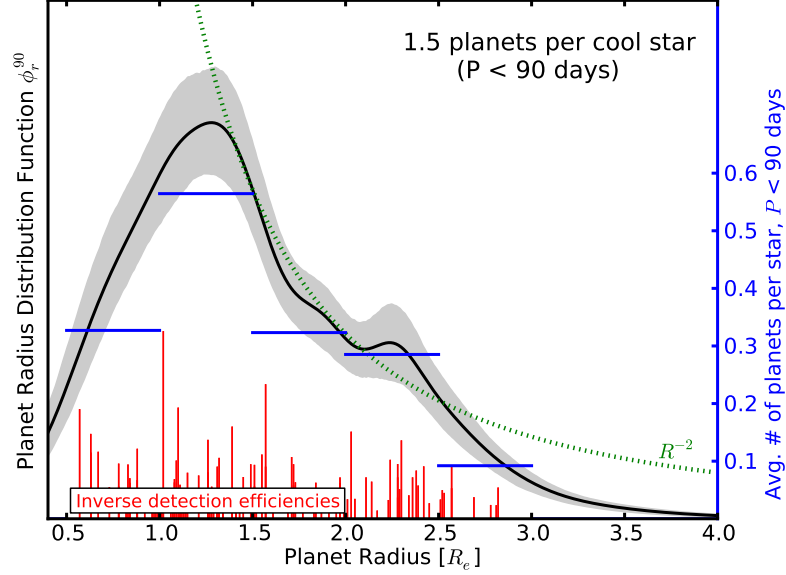


Figure 7.4 The empirical radius distribution of planets orbiting M dwarfs with periods  $<90$  days (black continuous curve), estimated with a weighted kernel density estimator (wKDE; see §7.3.3), with the bootstrap resampling-derived  $1\sigma$  uncertainty swath shaded grey—essentially a running Poisson error bar. The detection efficiency as a function of signal-to-noise ratio has been quantified by an SNR ramp from 0 at SNR = 6 to 1 at SNR = 12. The blue horizontal lines represent the standard “occurrence rate per bin” calculations for this sample. The vertical red lines represent the radii of individual planets in the sample, with their heights in proportion to the weight factors  $w_i$ . The green dotted curve is an  $R^{-2}$  power law, which bears a striking (and uncontrived) resemblance to the shape of this non-parametric radius function between about  $1.25$  and  $2 R_\oplus$ ; below about  $1.25 R_\oplus$ , the distribution appears to level off, and turn over below  $1 R_\oplus$ . Between about  $2$  and  $2.5 R_\oplus$  there appears to be an excess over a smooth distribution; this may be caused by a significant population of planets with H/He atmospheres. There is an average of 1.5 planets per cool star in orbits  $<90$  days over this radius range, and there is an average of greater than 0.5 planets per cool star in this period range for radii between  $1$  and  $1.5 R_\oplus$ .

### 7.3.3 Radius Distribution

Once  $\phi_{\text{SNR},i}$  is constructed for every planet, we then calculate  $\eta_{\text{disc},i}$  for each planet according to Equation 7.8, which combined with the transit probability gives the wKDE weight  $w_i$  for each planet, thus building  $\hat{\phi}_r^{90}$  (Equation 7.10). This function is plotted as the solid black line in Figure 7.4. To estimate the variance of this density estimator, we perform 1000 different bootstrap resamplings of the true planet dataset and recreate the wKDE for each resampling. The  $1\sigma$  uncertainty region determined by this procedure is illustrated as the grey shaded region in Figure 7.4, and is conceptually equivalent to a running Poisson error bar. The widths  $\sigma_i$  used to smooth the wKDE are taken to be twice the individual planet radius uncertainties, as this smooths out high-frequency wiggles while retaining broad features. Table 7.1 presents the data that goes into constructing

this distribution.

## 7.4 Results

The overall normalization of the radius function shown in Figure 7.4 indicates that there are approximately 1.5 planets per cool star with periods  $< 90$  d. In addition, there are several notable features of this distribution. The first is the peak between 1 and  $1.5 R_{\oplus}$  and the turnover below, both of which are robust features of the empirical distribution as characterized by the bootstrap uncertainty analysis. If this feature continues to hold as more and more candidates are identified around cool stars, it would point to a dramatic feature of planet formation and evolution:  $\sim 1 R_{\oplus}$  is the most common planet size to survive long-term in short orbits around cool stars. This might be understood by an explanation similar to that provided to explain the origins of the inner Solar System (Goldreich et al. 2004; Chambers 2001): a large number of isolation-mass protoplanets form quickly, and once the gas and planetesimals disk dissipates, a period of dynamical instability follows, at the end of which typically only a few larger planets remain, the rest having been either destroyed (or merged) via collisions or been swallowed by the host star. It is certainly plausible that  $\sim 1 R_{\oplus}$  planets might be the most likely outcome of this process, as this is precisely what has happened with the inner Solar System, with an outcome of two planets about the size of Earth.

The second notable feature of the distribution is the plateau between about 2 and  $2.5 R_{\oplus}$  and the steep decline above. This could plausibly be the effect of atmospheres, with planets massive enough to retain primordial H/He envelopes showing up as an excess population of planets this size compared to what would be expected from extrapolating toward larger radii from  $\sim 1$ – $1.5 R_{\oplus}$ . One prediction of this hypothesis would be that most of the planets smaller than  $2 R_{\oplus}$  are on average more dense than planets between 2 and  $3 R_{\oplus}$ , pointing to a smoother underlying mass distribution.

Finally, this distribution indicates that planets larger than  $\sim 3 R_{\oplus}$  are very rare around cool stars, consistent with the findings of RV surveys (Endl et al. 2003; Johnson et al. 2010b,a; Bonfils et al. 2013). There has been one hot Jupiter identified around a star in this sample (KOI-254b/Kepler-45b Johnson et al. 2012) and another recent discovery of note (Triaud et al. 2013), but such planets are clearly exceptional—the vast majority of close-in planets around cool stars are smaller than  $\sim 3 R_{\oplus}$ . Even Gliese 1214b (Charbonneau et al. 2009), by far the best-studied planet around an M dwarf to date, appears to be an exception to the typical system, as its radius of  $2.7 R_{\oplus}$  falls far down the tail of this distribution. In fact, there are  $\sim 30\times$  more planets smaller than Gl 1214b than there are larger than Gl 1214b—this bodes very well for the future of ground-based surveys, both transit and RV, as they can become more sensitive to smaller planets.

To explore the degree to which the SNR ramp and the period redistribution affect the shape of the derived radius function, we repeat this analysis using only the Q1-Q6 KOI catalog (Batalha et al. 2012) and out to a period of 50 days, to match with the analysis of Dressing & Charbonneau (2013). We then compare our full analysis to using a strict  $\text{SNR} = 7.1$  detection threshold (i.e.  $\eta_{\text{SNR}} = 1$  above  $\text{SNR} = 7.1$  and  $\eta_{\text{SNR}} = 0$



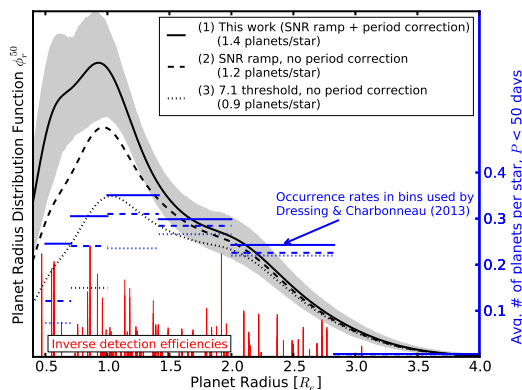


Figure 7.5 The planet radius distribution for  $P < 50$  days, using the Batalha et al. (2012) catalog in order to compare to previous studies, demonstrating the effect of the corrections accounted for in this work. The continuous curves are the non-parametric empirical density estimates, and the horizontal blue lines are the occurrence rates per bin, using the same bins as Dressing & Charbonneau (2013). The vertical red lines represent the radii of individual planets in the sample, with their heights in proportion to the weight factors  $w_i$ . The non-solid linestyles represent different analysis methods. Whereas Method 1 (solid lines) uses the full analysis described in this paper, with detection efficiency described by an SNR ramp (Fressin et al. 2013) and  $\phi_{\text{SNR},i}$  constructed by assigning random periods (§7.3.2), Methods 2 (dashed) and 3 (dotted) both keep period fixed when constructing  $\phi_{\text{SNR},i}$ , and Method 3 uses an SNR = 7.1 detection threshold rather than the SNR ramp. Method 2 is similar to the occurrence calculation in Fressin et al. (2013) and Method 3 uses the methods employed in Dressing & Charbonneau (2013). The importance of both a well-characterized detection efficiency function and treating period-based incompleteness correctly is clear: incorporating both these considerations significantly changes both the qualitative shape (especially as visualized with histograms) and normalization of the radius function below  $2 R_{\oplus}$ . In particular, the Dressing & Charbonneau (2013) analysis underestimates the occurrence rates of planets between  $0.5$  and  $1.4 R_{\oplus}$  by about a factor of two.

below), and also to using an alternative construction of  $\phi_{\text{SNR},i}$  where the period of planet  $i$  is kept fixed. Figure 7.5 illustrates these alternative estimates of the radius function. Method (2), illustrated by the dashed line, mirrors the analysis of Fressin et al. (2013), who corrected detections for the SNR ramp effect, but not for the period distribution. Method (3), illustrated by the dotted line, mirrors the analysis of Dressing & Charbonneau (2013), who used an SNR = 7.1 threshold and did not correct for orbital period. We show that these corrections make for a nearly 50% increase in the total inferred number of planets/star for  $P < 50$ d, and notably about a factor of two increase in the number of planets smaller than  $1.4 R_{\oplus}$ . From this comparison, we estimate that the true mean number of Earth-sized ( $0.5$ - $1.4 R_{\oplus}$ ) planets in the habitable zones (HZs) of these cool stars is at least twice as high as the lower limit estimated by Dressing & Charbonneau (2013); that is, probably closer to  $\sim 0.30$ , rather than  $0.15$ . Using revised calculations of the HZ, Kopparapu (2013) calculates a rate of  $\sim 0.50$  habitable Earth-sized planets around cool stars; the completeness considerations in this paper should increase that estimate to  $\sim 1$  planet/star.

Figures 7.4 and 7.5 also display the results of these calculations in the more traditional format of a histogram of planet occurrence in different radius bins. In order to calculate these histograms, we simply add up all the weights in each of the radius bins. Figure 7.4 uses linearly spaced bins; Figure 7.5 uses the same log-

Table 7.1. Data Used in Radius wKDE

KOI	$R_p [R_\oplus]$	$\sigma_R$	$p_{\text{tr}}$	$\eta_{\text{disc}}$	$w_i$
KOI961.03 <sup>a</sup>	0.57	0.18	0.051	0.14	140.1
KOI2453.01 <sup>a</sup>	0.63	0.11	0.077	0.12	108.2
KOI2542.01 <sup>a</sup>	0.63	0.08	0.132	0.09	84.2
KOI1422.03 <sup>a</sup>	0.67	0.11	0.051	0.23	85.3
KOI961.02 <sup>a</sup>	0.73	0.20	0.132	0.19	39.9
KOI251.02 <sup>b</sup>	0.78	0.09	0.045	0.55	40.4
KOI961.01 <sup>a</sup>	0.78	0.22	0.068	0.21	70.0
KOI952.05 <sup>a</sup>	0.82	0.08	0.180	0.44	12.6
KOI1843.02 <sup>a</sup>	0.83	0.12	0.044	0.33	68.9
KOI2006.01 <sup>a</sup>	0.84	0.07	0.067	0.33	45.2
KOI2238.01 <sup>a</sup>	0.84	0.13	0.093	0.21	51.2
KOI1146.01 <sup>a</sup>	0.88	0.18	0.034	0.33	89.1
KOI1702.01 <sup>a</sup>	0.88	0.13	0.076	0.27	48.7
KOI250.03 <sup>b</sup>	0.88	0.19	0.059	0.69	24.6
KOI2662.01 <sup>a</sup>	0.92	0.22	0.085	0.33	35.7
KOI2036.02 <sup>a</sup>	0.96	0.16	0.047	0.56	38.0
KOI2306.01 <sup>a</sup>	0.97	0.08	0.241	0.51	8.1
KOI1649.01 <sup>b</sup>	0.98	0.12	0.057	0.52	33.7
:	:	:	:	:	:
KOI255.01 <sup>a</sup>	2.57	0.09	0.016	0.92	67.9
KOI2926.01 <sup>b</sup>	2.57	0.29	0.029	0.90	38.3
KOI248.01 <sup>b</sup>	2.69	0.31	0.041	0.91	26.8
KOI531.01 <sup>b</sup>	2.78	0.40	0.062	0.96	16.8
KOI2156.01 <sup>a</sup>	2.81	0.20	0.068	0.89	16.5
KOI781.01 <sup>a</sup>	2.82	0.10	0.028	0.91	39.2

<sup>a</sup>Planet radius based on spectroscopic stellar parameters from the analysis of Muirhead et al. (2012a)

<sup>b</sup>Spectroscopic stellar characterization not available, or  $T_{\text{eff}} > 3800$ , so planet radius based on stellar parameters from Dressing & Charbonneau (2013)

arithmetic bins used by Dressing & Charbonneau (2013) for comparison. There are several qualitative points to note regarding these histograms. The first is that they can be visually deceptive: for example, the dotted histogram in Figure 7.5 is approximately flat between 1 and 2  $R_\oplus$  before decreasing in the 0.7-1  $R_\oplus$  bin, even though the smoothed distributions continue rising steadily all the way to  $\sim 1.25 R_\oplus$ —a result of logarithmically spaced bins. Secondly, they only provide a very coarse description of the shape of the radius distribution; that is, there is much more detectable structure than can be captured in a few bins—for example, the plateau in Figure 7.4 between 2 and 2.5  $R_\oplus$  is not captured in the histogram illustration, nor is the steep fall-off above 2.5  $R_\oplus$ , nor the identification of  $\sim 1.25 R_\oplus$  as the location of the low-end turnover.

## 7.5 Exploring Assumptions

There are two assumptions that we have made to construct this radius distribution:

- The period distribution of planets is both independent of planet radius and well-characterized by the current planet detections.

- The detection efficiency of planets in the Batalha et al. (2012) catalog follows an SNR ramp similar to that described in Fressin et al. (2013).

Are these assumptions justified? What are the implications if they are incorrect?

### 7.5.1 Period Distribution Assumption

Figure 7.1 illustrates very clearly why the detected population of small planets in the Cool KOI sample is indeed very likely incomplete, showing that where the detected period distribution of the smallest of the Cool KOIs drops off is right around the periods where the known short-period small KOIs would have become undetectable. This is the motivation behind the period redistribution procedure we use to calculate  $\eta_{\text{disc},i}$  in §7.3—correcting for the undetectable longer-period small planets. Such a correction is surely needed; however, the nature of this correction as applied in this work—using the implied all-planet period distribution for each planet—merits some discussion.

There are certainly both physical reasons and observational suggestions to believe that the planet period distribution is *not* completely independent of radius. In particular, Howard et al. (2012) finds (shown in their Figure 6) that the fraction of short-period planets that are large ( $4\text{--}8 R_{\oplus}$ ) is smaller than the fraction of longer-period planets that are large; in other words, the period distribution of larger planets decreases (heading towards shorter periods) sooner than does the distribution of smaller planets ( $2\text{--}4 R_{\oplus}$ ). Dong & Zhu (2012) present a similar finding. While there is not yet compelling evidence that this same effect has been detected for planets smaller than  $2 R_{\oplus}$ , simple physical considerations such as increasing stellar insolation (Weiss & Marcy 2013) might reasonably contribute to a dearth of larger planets on short-period orbits. However, as there is no corresponding clear physical explanation for the absence of smaller planets in longer orbits, it is reasonable to assume that they do in fact exist, and that their period distribution might resemble the period distribution of the larger planets that are detected in such orbits.

In addition, if small planets are in general more common than larger planets (as it appears), and small planets are not being detected on longer periods, then approximating the distribution of all planets with just the total observed distribution will naturally *underestimate the total numbers of longer-period planets*. In fact, looking at the all-planet distribution in Figure 7.2, it is quite reasonable to expect that perhaps the apparent decrease of the distribution function longer than  $\sim 20$  days is actually due to the fact that only planets larger than  $1.5 R_{\oplus}$  or so (which may very well be a minority of all planets) are being readily detected at these periods. And so, this will cause an overestimate of the true fraction of large planets that are at shorter periods, and an underestimate of the true numbers of small planets that are at longer periods.

The total effect of this assumption will thus be to systematically shift the simulated “planet clone” distributions towards shorter periods, and thus the SNR distributions  $\phi_{\text{SNR},i}$  toward larger SNRs. This will cause  $\eta_{\text{disc},i}$  values to be slightly overestimated, which will lead to underestimating the weights  $w_i$  and subsequently the normalization of the radius function, especially towards smaller planets, which depend most heavily on

this correction. However, this effect is a very small one—we repeated the analysis using a distribution in  $\log P$  that has an exponential cutoff below 10 days and is flat above (that is, without the dip beyond  $\sim 15$  days that appears in Figure 7.2), and there is negligible difference in the ensuing radius distribution, with the only change being a few percent increase in the planet occurrence in the  $1\text{--}1.5 R_{\oplus}$  bin in Figure 7.4—and notably no change to the  $0.5\text{--}1 R_{\oplus}$  bin.

### 7.5.2 SNR Ramp Assumption

Figure 7.5 shows that quantifying the detection efficiency of the *Kepler* pipeline as an SNR ramp (Equation 7.7) rather than a strict threshold cut makes a significant difference in the inferred occurrence rate of planets smaller than  $2 R_{\oplus}$ . In particular, the value of SNR by which the pipeline is assumed to be 100% complete changes the overall normalization of the low end of the radius distribution. While the 6-16 SNR ramp characterization of the efficiency for the Batalha et al. (2012) catalog was introduced and defended by Fressin et al. (2013), and was adjusted in this work (Figure 7.4) to be relevant to the Q1-Q8 catalog (F. Fressin, priv. comm.), it should be treated as a temporary solution until the detection efficiency of the *Kepler* pipeline can be directly quantified as a function of SNR through injection/recovery simulations. However, the notable features in Figures 7.4 and 7.5—the rise down to about  $1 R_{\oplus}$  and the small plateau around  $2.5 R_{\oplus}$ —are robust to the precise details of the ramp.

## 7.6 Conclusions

We apply a simple non-parametric method of analyzing the empirical shape of the planet radius distribution from a transit survey—the weighted kernel density estimator, or wKDE. This estimator is similar to a standard kernel density estimator, except that its overall normalization is constructed to be equal to the total number of planets per star, and that each data point is weighted according to its inverse detection efficiency. We also show that properly computing this efficiency requires two considerations that have not always been applied in previous occurrence rate studies: correcting for the planet period distribution when calculating how many target stars around which a particular planet could have been observed, and considering that the detection efficiency is a rising function of signal-to-noise ratio, and not just a strict cutoff.

Applying this analysis to the 113 planet candidates currently in the cumulative KOI catalog with periods less than 90 days discovered around the cool *Kepler* targets photometrically characterized by Dressing & Charbonneau (2013), we identify several key features of the radius distribution of small planets around small stars which invite theoretical explanation. First, even correcting carefully for incompleteness, the data indicate a flattening or turnover of the distribution around about  $1\text{--}1.25 R_{\oplus}$ , suggesting that planets about this size are the most common to survive in short orbits around cool stars. Notably, this feature of the distribution is robust to incompleteness below  $1 R_{\oplus}$ : it appears that planets smaller than Earth are indeed more rare around cool stars than planets around the size of Earth. Secondly, there appears to be a plateau from about 2 to  $2.5 R_{\oplus}$

$R_{\oplus}$ , where there is an overabundance of planets as compared to what would be predicted from a smooth distribution extrapolating from small to larger planet radii, perhaps an indication of a population of planets with significant H/He atmospheres. And finally, the occurrence pattern of planets around cool stars indicates that there are many planets just beyond the detection threshold of ground-based surveys, as planets larger than Gl 1214b ( $2.7 R_{\oplus}$ ) are  $\sim 30\times$  rarer than planets with  $R_p < 2.7 R_{\oplus}$ .

Comparing this non-parametric radius function estimate with the more traditional presentation of planet occurrence rate in different radius bins demonstrates that the bin presentation can be visually misleading, in addition to missing details of the distribution that are accessible in the data. And comparing our results to the occurrence calculations of Dressing & Charbonneau (2013), we find that there are about a factor of two more planets from  $0.5$  to  $1.4 R_{\oplus}$  than that analysis determined; this would imply that there are an average of  $\sim 0.30$  habitable-zone Earth-like planets per cool star, rather than the  $\sim 0.15$  estimated by that work. And if this same correction is made to the calculations of Kopparapu (2013), which use updated HZ calculations but the same occurrence formalism as Dressing & Charbonneau (2013), then this number would become closer to  $\sim 1$  planet per star. Habitable-zone, Earth-sized planets abound throughout the Galaxy in numbers even larger than previously estimated.

In addition to demonstrating how to extract empirical distributions from *Kepler* data without relying on arbitrary binning, we call attention to the importance of understanding in detail the detection efficiency of transit search algorithms. Future studies can most directly support analyses such as these—which lie at the very core of the *Kepler* mission—by directly computing the detection efficiency of these algorithms as a function of signal-to-noise ratio, enabling reliable correction for incompleteness near the detection threshold, where many of the most scientifically interesting discoveries will be made. And finally, we emphasize that this calculation is based on a target sample of only about 3900 cool stars and a KOI search only through Q8 data. Continued expansion of the cool star *Kepler* sample by re-appropriation of target pixels could potentially increase this sample size by a factor of two or more, allowing for greatly strengthened conclusions from the small-planet radius distribution and giving a greater handle on the formation processes of planetary systems around the most numerous stars in the Galaxy. In addition, careful application of these same principles to the entire *Kepler* dataset, as permitted by accurate knowledge of stellar parameters, will continue to uncover important clues to the formation and evolution of all types of planetary systems.

## Chapter 8

# Conclusion

In Chapter 1, I reviewed what was known about population statistics of exoplanets before the launch of the *Kepler* mission. In Chapters 4, 5, 6, and 7, I detailed the contributions that I have made to further understanding exoplanet demographics and enabling confidence in large demographic studies that treat *Kepler* candidates as if they were true planets. In this brief conclusion, I summarize the most important of these studies and look forward to what I see as the future of exoplanet demographics.

The two most readily available physical properties of a planet candidate are its period and radius; consequently, the most natural investigation into the *Kepler* candidates is to try to reconstruct the joint period-radius distribution of exoplanets. Soon after the release of the first candidate catalog (Borucki et al. 2011), two independent analyses were released. Howard et al. (2012, H12) focused on the sample of bright (*Kepler* magnitude  $< 15$ ) “Solar-type” stars ( $T_{\text{eff}} = 4100\text{--}6100$  K) and planet candidates with  $P < 50$  days, in order to be able to compare to the RV results of the *NASA-UC Eta-Earth* survey. The analysis also was restricted to planet radii  $> 2 R_{\oplus}$  and  $\text{SNR} > 10$ —this constraint and the *Kepler* magnitude cut were intended to simplify analysis by choosing candidates whose detection efficiencies should be close to 100% (i.e. avoiding having to deal with the kinds of corrections that I presented in Chapter 7). Using this subset, they analyzed planet

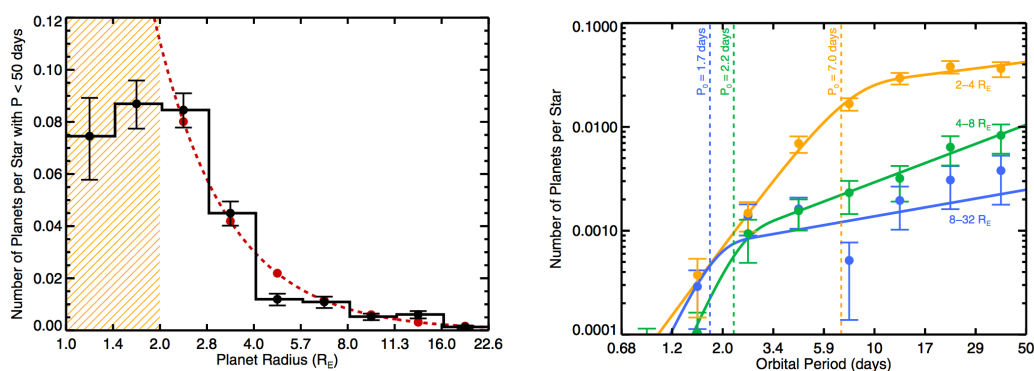


Figure 8.1 The radius and period distributions of planets around solar-type stars (4100–6100 K), derived from the distribution of *Kepler* planet candidates by Howard et al. (2012). These results depended on assuming that all the candidates were planets; i.e. that the false positive rate was negligible.

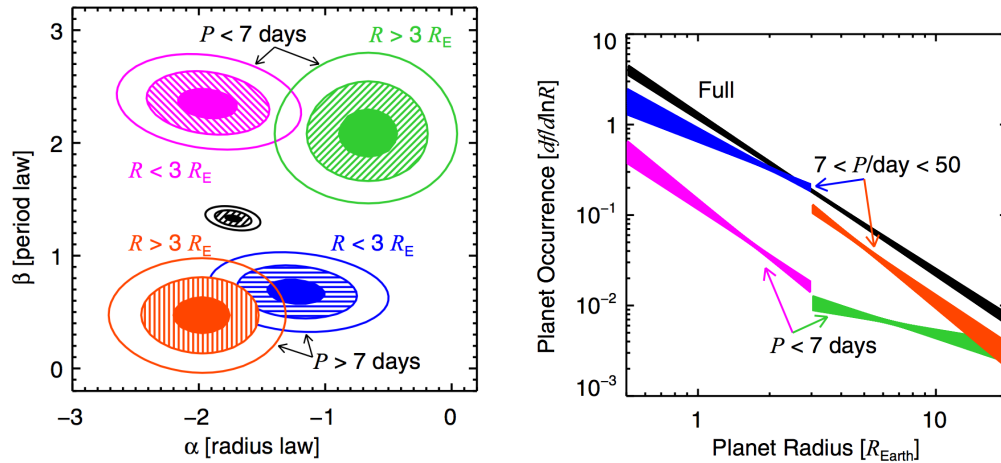


Figure 8.2 Summary of the derivation of the underlying shape of the exoplanet joint period-radius distribution, assuming a joint power-law distribution  $\propto R^\alpha P^\beta$  (Youdin 2011). The population of planet candidates from Borucki et al. (2011) is split into four quadrants, short/long-period (split at 7 days) and small/large (split at  $3 R_\oplus$ ). The most striking feature found here is the qualitatively different shape of the radius function for short and long period planets. In addition (not indicated in these plots), extrapolation of these power-law fits indicates that there could be an average of  $\approx 3$  Earth-sized planets per star out to 365-day orbits.

occurrence in bins of radius and period, correcting for transit geometry and detection completeness. Figure 8 displays two different views into this distribution: the overall radius distribution and the period distribution for candidates of different sizes. One of the most notable discoveries in this work, is the finding that the period distributions of planets, as parametrized by a power law + exponential fit, have different cutoff periods depending on planet size (right hand panel of Figure 8).

Youdin (2011) took a slightly different strategy than H12, laying out an analytic formalism to fit an underlying radius-period distribution  $\propto R^\alpha P^\beta$  to the observed data, accounting for detection efficiency and transit geometry. Rather than conservatively stopping at  $R_p = 2 R_\oplus$  as did H12, this analysis extrapolated the survey detection efficiency down to planet candidates of  $0.5 R_\oplus$ . Figure 8 summarizes the results of fitting this joint radius-period distribution to four “quadrants” of candidates from Borucki et al. (2011), split at a period of 7 days and radius of  $3 R_\oplus$ . The finding, similar to H12 is that there appears to be a break in the period distribution at short periods, and also that the size distributions of shorter and longer-period planets are qualitatively different (RH panel of Figure 8). In addition to these power-law fits, Youdin (2011) also calculates a non-parametric occurrence rate of an average of about 0.7-1.4 planets per star  $> 0.5 R_\oplus$  and  $P < 50$  days (the range depending on systematic uncertainties in the *Kepler* discovery efficiency). And finally, a speculative extrapolation of the power-law fits out to periods of 1 year implies an astonishing average of  $\approx 3$  planets per star larger than Earth with periods less than 365 days.

Complementary to these investigations of the occurrence of planets in the mass-radius plane has been what *Kepler* has taught us about multi-planet systems. The Batalha et al. (2012) catalog contained 885 planet candidates in 361 multi-planet systems. This large number of multi-transiting systems has enabled studies of

planetary system mutual inclination, all of which indicate that multiple transiting planets tend to have orbits coplanar to within a few degrees (Fabrycky et al. 2012b; Fang & Margot 2012). In addition, the population of detections seems to indicate that 75%-80% of planetary systems have one or two planets with orbital periods less than 200 days (Fang & Margot 2012).

While all these studies have shown for the first time the broad characteristics of exoplanet populations at relatively short periods down to approximately the size of Earth, further analysis of *Kepler* data will continue to reveal still more. In particular, I do not believe that further studying planet “occurrence rates” or statistics like the average number of planets per star or the fraction of stars with planets is the most fruitful direction for future study of exoplanet demographics, nor any other studies that make any assumptions that planetary systems are able to be characterized as a single population. In fact, I believe what the *Kepler* results are hinting at is that, indeed, the fraction of stars with planets is 1—in other words, all stars probably host planetary systems.

I think the direction that the most illuminating future exoplanet demographic studies will take is answering questions more like “What are the different *kinds* of planetary systems, and what are their relative frequency?” “What are the different factors that cause different planetary systems to end up in different end states?” or “What role does stellar binarity or other environmental factors play in planetary system formation and evolution?” Answering these questions will doubtless require careful joint analysis of several data sets: *Kepler* candidate catalogs, *Kepler* follow-up data (especially spectroscopy and imaging to characterize the physical properties and binarity statistics of *Kepler* host stars), and the ensemble of radial velocity survey data (which has been somewhat tabled in recent years due to the frenzy regarding *Kepler* results). Eventual success in this endeavor, however, will finally enable us to address the fundamental questions that motivate much of exoplanet research: how typical is our Solar System, and how common are planetary systems that might be able to support life?



# Bibliography

- Almenara, J. M., Deeg, H. J., Aigrain, S., et al. 2009, *A&A*, 506, 337
- Alonso, R., Brown, T. M., Torres, G., et al. 2004, *ApJ*, 613, L153
- Baglin, A. 2003, *Advances in Space Research*, 31, 345
- Bakos, G. Á., Noyes, R. W., Kovács, G., et al. 2007, *ApJ*, 656, 552
- Ballard, S., Fabrycky, D., Fressin, F., et al. 2011, *ApJ*, 743, 200
- Baraffe, I., Chabrier, G., Allard, F., & Hauschildt, P. H. 2002, *A&A*, 382, 563
- Barclay, T., Rowe, J. F., Lissauer, J. J., et al. 2013, *Nature*, 494, 452
- Basri, G., Walkowicz, L. M., Batalha, N., et al. 2010, *ApJ*, 713, L155
- Batalha, N. M., Borucki, W. J., Bryson, S. T., et al. 2011, *ApJ*, 729, 27
- Batalha, N. M., Borucki, W. J., Koch, D. G., et al. 2010a, *ApJ*, 713, L109
- Batalha, N. M., Rowe, J. F., Bryson, S. T., et al. 2012, *ArXiv e-prints*
- Batalha, N. M., Rowe, J. F., Gilliland, R. L., et al. 2010b, *ApJ*, 713, L103
- Batygin, K. & Morbidelli, A. 2013, *AJ*, 145, 1
- Batygin, K. & Stevenson, D. J. 2010, *ApJ*, 714, L238
- Beatty, T. G. & Gaudi, B. S. 2008a, *ApJ*, 686, 1302
- Beatty, T. G. & Gaudi, B. S. 2008b, *ApJS*, 686, 1302
- Biller, B. A., Close, L. M., Masciadri, E., et al. 2007, *ApJS*, 173, 143
- Bodenheimer, P., Laughlin, G., & Lin, D. N. C. 2003, *ApJ*, 592, 555
- Bonfils, X., Delfosse, X., Udry, S., et al. 2013, *A&A*, 549, A109
- Borucki, W., Koch, D., Basri, G., et al. 2008, in *IAU Symposium*, Vol. 249, *IAU Symposium*, ed. Y.-S. Sun, S. Ferraz-Mello, & J.-L. Zhou, 17–24

- Borucki, W. J., Koch, D. G., Basri, G., et al. 2011, ArXiv e-prints
- Borucki, W. J., Koch, D. G., Batalha, N., et al. 2012, ApJS, 745, 120
- Boss, A. P. 1997, Science, 276, 1836
- Bouchy, F., Bonomo, A. S., Santerne, A., et al. 2011, A&A, 533, A83
- Bouchy, F., Udry, S., Mayor, M., et al. 2005, A&A, 444, L15
- Brown, T. M. 2003, ApJ, 593, L125
- Burke, C. 2013, in prep
- Burrows, A., Hubeny, I., Budaj, J., & Hubbard, W. B. 2007, ApJ, 661, 502
- Butler, R. P., Vogt, S. S., Marcy, G. W., et al. 2004, ApJ, 617, 580
- Cassan, A., Kubas, D., Beaulieu, J.-P., et al. 2012, Nature, 481, 167
- Chabrier, G. 2001, ApJ, 554, 1274
- Chabrier, G., Baraffe, I., Allard, F., & Hauschildt, P. 2000, ApJ, 542, 464
- Chambers, J. E. 2001, , 152, 205
- Charbonneau, D., Berta, Z. K., Irwin, J., et al. 2009, Nature, 462, 891
- Charbonneau, D., Brown, T. M., Latham, D. W., & Mayor, M. 2000, ApJ, 529, L45
- Chatterjee, S., Ford, E. B., Matsumura, S., & Rasio, F. A. 2008, ApJ, 686, 580
- Chiang, E. & Laughlin, G. 2012, ArXiv e-prints
- Claret, A. 2000, A&A, 363, 1081
- Collier Cameron, A., Bouchy, F., Hébrard, G., et al. 2007, MNRAS, 375, 951
- Colón, K. D., Ford, E. B., & Morehead, R. C. 2012, MNRAS, 426, 342
- Cumming, A., Butler, R. P., Marcy, G. W., et al. 2008, PASP, 120, 531
- Dong, S. & Zhu, Z. 2012, ArXiv e-prints
- Dotter, A., Chaboyer, B., Jevremović, D., et al. 2008, ApJS, 178, 89
- Doyle, L. R., Carter, J. A., Fabrycky, D. C., et al. 2011, Science, 333, 1602
- Dressing, C. D. & Charbonneau, D. 2013, ArXiv e-prints

- Endl, M., Cochran, W. D., Tull, R. G., & MacQueen, P. J. 2003, *AJ*, 126, 3099
- Evans, T. M. & Sackett, P. D. 2010, *ApJ*, 712, 38
- Fabrycky, D. & Tremaine, S. 2007, *ApJ*, 669, 1298
- Fabrycky, D. C., Ford, E. B., Steffen, J. H., et al. 2012a, *ApJ*, 750, 114
- Fabrycky, D. C., Lissauer, J. J., Ragozzine, D., et al. 2012b, *ArXiv e-prints*
- Fabrycky, D. C. & Winn, J. N. 2009, *ApJ*, 696, 1230
- Fang, J. & Margot, J.-L. 2012, *ApJS*, 761, 92
- Feroz, F., Hobson, M. P., & Bridges, M. 2009, *MNRAS*, 398, 1601
- Fischer, D. A. & Valenti, J. 2005, *ApJ*, 622, 1102
- Ford, E. B. & Rasio, F. A. 2008, *ApJ*, 686, 621
- Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2012, *ArXiv e-prints*
- Fortney, J. J., Marley, M. S., & Barnes, J. W. 2007, *ApJ*, 659, 1661
- Fressin, F., Torres, G., Charbonneau, D., et al. 2013, *ArXiv e-prints*
- Fressin, F., Torres, G., Désert, J.-M., et al. 2011, *ApJS*, 197, 5
- Fressin, F., Torres, G., Rowe, J. F., et al. 2012, *Nature*, 482, 195
- Gaudi, B. S. 2005, *ApJ*, 628, L73
- Gaudi, B. S., Seager, S., & Mallen-Ornelas, G. 2005, *ApJ*, 623, 472
- Gaudi, B. S. & Winn, J. N. 2007, *ApJ*, 655, 550
- Girardi, L., Bertelli, G., Bressan, A., et al. 2002, *A&A*, 391, 195
- Girardi, L., Groenewegen, M. A. T., Hatziminaoglou, E., & da Costa, L. 2005, *A&A*, 436, 895
- Goldreich, P., Lithwick, Y., & Sari, R. 2004, *ARA&A*, 42, 549
- Gould, A., Pepper, J., & DePoy, D. L. 2003, *ApJS*, 594, 533
- Hansen, B. & Murray, N. 2013, *ArXiv e-prints*
- Hansen, B. M. S. & Murray, N. 2012, *ApJ*, 751, 158
- Hartman, J. D., Bakos, G., Stanek, K. Z., & Noyes, R. W. 2004, *AJ*, 128, 1761

- Hébrard, G., Bouchy, F., Pont, F., et al. 2008, *A&A*, 488, 763
- Henry, G. W., Marcy, G. W., Butler, R. P., & Vogt, S. S. 2000, *ApJ*, 529, L41
- Ho, S. & Turner, E. L. 2010a, *ArXiv e-prints*
- Ho, S. & Turner, E. L. 2010b, *ArXiv e-prints*
- Holman, M. J., Fabrycky, D. C., Ragozzine, D., et al. 2010, *Science*, 330, 51
- Horne, K. 2003, in *Astronomical Society of the Pacific Conference Series*, Vol. 294, *Scientific Frontiers in Research on Extrasolar Planets*, ed. D. Deming & S. Seager, 361–370
- Howard, A. W., Marcy, G. W., Bryson, S. T., et al. 2012, *ApJS*, 201, 15
- Howard, A. W., Marcy, G. W., Johnson, J. A., et al. 2010, *Science*, 330, 653
- Jenkins, J. M., Borucki, W. J., Koch, D. G., et al. 2010a, *ApJ*, 724, 1108
- Jenkins, J. M., Caldwell, D. A., Chandrasekaran, H., et al. 2010b, *ApJ*, 713, L120
- Johnson, J. A., Aller, K. M., Howard, A. W., & Crepp, J. R. 2010a, *PASP*, 122, 905
- Johnson, J. A. & Apps, K. 2009, *ApJ*, 699, 933
- Johnson, J. A., Gazak, J. Z., Apps, K., et al. 2012, *AJ*, 143, 111
- Johnson, J. A., Howard, A. W., Marcy, G. W., et al. 2010b, *PASP*, 122, 149
- Johnson, J. A., Winn, J. N., Albrecht, S., et al. 2009, *PASP*, 121, 1104
- Jurić, M. & Tremaine, S. 2008, *ApJ*, 686, 603
- Koch, D. G., Borucki, W., Webster, L., et al. 1998, in *Presented at the Society of Photo-Optical Instrumentation Engineers (SPIE) Conference*, Vol. 3356, *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, ed. P. Y. Bely & J. B. Breckinridge, 599–607
- Konacki, M., Torres, G., Sasselov, D. D., & Jha, S. 2003, *ApJ*, 597, 1076
- Kopparapu, R. 2013, *arxiv:1303.2649*
- Kozai, Y. 1962, *AJ*, 67, 591
- Lai, D., Foucart, F., & Lin, D. N. C. 2010, *ArXiv e-prints*
- Latham, D. W., Bakos, G. Á., Torres, G., et al. 2009, *ApJ*, 704, 1107
- Latham, D. W., Brown, T. M., Monet, D. G., et al. 2005, in *Bulletin of the American Astronomical Society*, Vol. 37, *Bulletin of the American Astronomical Society*, 1340–+

- Lin, D. N. C., Bodenheimer, P., & Richardson, D. C. 1996, *Nature*, 380, 606
- Lissauer, J. J., Fabrycky, D. C., Ford, E. B., et al. 2011, *Nature*, 470, 53
- Lithwick, Y. & Wu, Y. 2012, *ApJ*, 756, L11
- Lithwick, Y., Xie, J., & Wu, Y. 2012, *ApJS*, 761, 122
- Lopez, E. D., Fortney, J. J., & Miller, N. 2012, *ApJS*, 761, 59
- Luhman, K. L. & Jayawardhana, R. 2002, *ApJ*, 566, 1132
- Mandel, K. & Agol, E. 2002, *ApJ*, 580, L171
- Mann, A. W., Gaidos, E., Lépine, S., & Hilton, E. J. 2012, *ApJ*, 753, 90
- Marcy, G. W. & Butler, R. P. 1996, *ApJ*, 464, L147+
- Matsumura, S., Peale, S. J., & Rasio, F. A. 2010, *ApJ*, 725, 1995
- Mayor, M. & Queloz, D. 1995, *Nature*, 378, 355
- McCullough, P. R., Stys, J. E., Valenti, J. A., et al. 2006, *ApJ*, 648, 1228
- Mizuno, H. 1980, *Progress of Theoretical Physics*, 64, 544
- Morton, T. D. 2012, *ApJ*, 761, 6
- Morton, T. D. & Johnson, J. A. 2011a, *ApJ*, 738, 170
- Morton, T. D. & Johnson, J. A. 2011b, *ApJ*, 738, 170
- Muirhead, P. S., Hamren, K., Schlawin, E., et al. 2012a, *ApJ*, 750, L37
- Muirhead, P. S., Johnson, J. A., Apps, K., et al. 2012b, *ApJ*, 747, 144
- Nagasawa, M., Ida, S., & Bessho, T. 2008, *ApJ*, 678, 498
- Narita, N., Enya, K., Sato, B., et al. 2007, *PASJ*, 59, 763
- Narita, N., Hirano, T., Sanchis-Ojeda, R., et al. 2010, *PASJ*, 62, L61+
- Narita, N., Sato, B., Hirano, T., & Tamura, M. 2009, *PASJ*, 61, L35
- O'Donovan, F. T., Charbonneau, D., Torres, G., et al. 2006, *ApJ*, 644, 1237
- Pepper, J., Gould, A., & Depoy, D. L. 2003, , 53, 213
- Petigura, E. & Marcy, G. 2013, in prep

- Poleski, R., McCullough, P. R., Valenti, J. A., et al. 2010, *ApJS*, 189, 134
- Pollack, J. B., Hubickyj, O., Bodenheimer, P., et al. 1996, , 124, 62
- Raghavan, D., McAlister, H. A., Henry, T. J., et al. 2010, *ApJS*, 190, 1
- Rein, H. 2012, *MNRAS*, 427, L21
- Rogers, L. A., Bodenheimer, P., Lissauer, J. J., & Seager, S. 2011, *ApJS*, 738, 59
- Santerne, A., Díaz, R. F., Moutou, C., et al. 2012, *ArXiv e-prints*
- Santos, N. C., Israelian, G., & Mayor, M. 2004, *A&A*, 415, 1153
- Sato, B., Fischer, D. A., Henry, G. W., et al. 2005, *ApJ*, 633, 465
- Schlaufman, K. C. 2010, *ApJ*, 719, 602
- Seager, S., Kuchner, M., Hier-Majumder, C. A., & Militzer, B. 2007, *ApJ*, 669, 1279
- Seager, S. & Mallén-Ornelas, G. 2003, *ApJ*, 585, 1038
- Steffen, J. H., Fabrycky, D. C., Agol, E., et al. 2013, *MNRAS*, 428, 1077
- Swift, J. J. & Beaumont, C. N. 2010, *PASP*, 122, 224
- Swift, J. J., Johnson, J. A., Morton, T. D., et al. 2013, *ApJS*, 764, 105
- Tokovinin, A., Thomas, S., Sterzik, M., & Udry, S. 2006, *A&A*, 450, 681
- Tokovinin, A. A. 1999, *VizieR Online Data Catalog*, 412, 40075
- Torres, G., Fressin, F., Batalha, N. M., et al. 2011, *ApJ*, 727, 24
- Torres, G., Konacki, M., Sasselov, D. D., & Jha, S. 2004, *ApJ*, 614, 979
- Torres, G., Konacki, M., Sasselov, D. D., & Jha, S. 2005, *ApJ*, 619, 558
- Triaud, A. H. M. J., Anderson, D. R., Collier Cameron, A., et al. 2013, *A&A*, 551, A80
- Triaud, A. H. M. J., Collier Cameron, A., Queloz, D., et al. 2010, *A&A*, 524, A25+
- Wang, B. & Wang, X. 2007, *ArXiv e-prints*
- Watson, C. A., Littlefair, S. P., Diamond, C., et al. 2010, *ArXiv e-prints*
- Weiss, L. & Marcy, G. 2013, in prep
- Winn, J. N., Fabrycky, D., Albrecht, S., & Johnson, J. A. 2010, *ApJ*, 718, L145

- Winn, J. N., Holman, M. J., Henry, G. W., et al. 2007, *AJ*, 133, 1828
- Winn, J. N., Johnson, J. A., Albrecht, S., et al. 2009a, *ApJ*, 703, L99
- Winn, J. N., Johnson, J. A., Fabrycky, D., et al. 2009b, *arxiv:0902.3461*
- Winn, J. N., Johnson, J. A., Marcy, G. W., et al. 2006, *ApJ*, 653, L69
- Winn, J. N., Noyes, R. W., Holman, M. J., et al. 2005, *ApJ*, 631, 1215
- Wolf, A. S., Laughlin, G., Henry, G. W., et al. 2007, *ApJ*, 667, 549
- Wolszczan, A. & Frail, D. A. 1992, *Nature*, 355, 145
- Wright, J. T. 2010, in *EAS Publications Series*, Vol. 42, *EAS Publications Series*, ed. K. Goździewski, A. Niedzielski, & J. Schneider, 3–17
- Wright, J. T., Fakhouri, O., Marcy, G. W., et al. 2011, *PASP*, 123, 412
- Wright, J. T., Marcy, G. W., Howard, A. W., et al. 2012, *ArXiv e-prints*
- Wu, Y. & Lithwick, Y. 2012, *ArXiv e-prints*
- Wu, Y., Murray, N. W., & Ramsahai, J. M. 2007, *ApJ*, 670, 820
- Youdin, A. N. 2011, *ApJ*, 742, 38